CHAPTER 1
THE MILKY WAY – OUR GALAXY

1.1 Introduction

To the observer, the Milky Way is the faint band of diffuse light that arches across the night sky from horizon to horizon (Figure 1.1). This light comes mainly from a multitude of stars, although the unaided eye is unable to resolve these stars individually, hence the appearance of a 'band' of light. The stellar nature of the visible Milky Way was revealed about four hundred years ago when Galileo Galilei (1564–1642) made some of the earliest astronomical observations using a telescope.

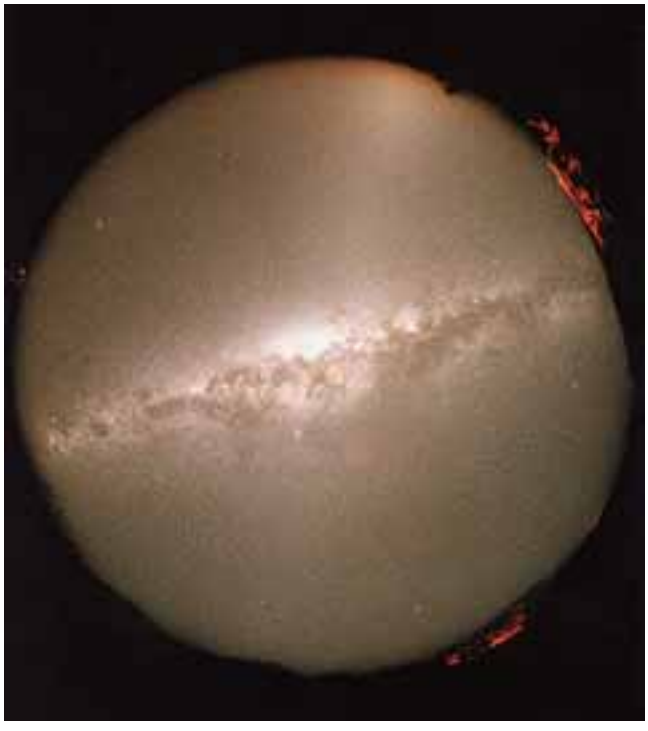


Figure 1.1 A photograph of one hemisphere of the night sky, showing the Milky Way stretching from horizon to horizon. The most prominent, central portion of the Milky Way is directly overhead in this image. The light comes mainly from the enormous number of stars that exist within our Galaxy; there are about 10^{11} in all, but many are too faint to see, or are obscured by the dust that is also part of the Galaxy. (D. di Cicco, Sky Publishing Corp.)

Although Galileo recognized the existence of huge numbers of stars in the Milky Way in around 1610, it was not until the 20th century that astronomers were able to deduce the distribution of those stars with any accuracy, and only over the past decade or two have they arrived at what is believed to be a true understanding of the nature of the Milky Way. We now know roughly how far the Milky Way extends in each direction in space, and that its light comes mostly from stars distributed in a flattened disc-like structure some 100 000 light-years (ly) across. We also know that, in addition to stars, the Milky Way contains gas and dust. Perhaps most astonishingly of all, the majority of astronomers have now become convinced that such familiar entities as stars, gas and dust account for no more than about 10% of the mass of the Milky Way. Most of the mass, a staggering 10^{12} times the mass of the Sun (i.e. $10^{12}M_{\odot}$, where $M_{\odot} \approx 2 \times 10^{30}$ kg), is now believed to be attributable to some kind of unidentified form of matter known, somewhat enigmatically, as 'dark matter'. The term **Milky Way** is now applied to this whole collection of entities – the system of stars, gas, dust, and dark matter – not just to the diffuse band of starlight resolved by Galileo.

Does anything exist beyond the boundary of the Milky Way? The answer is an emphatic 'yes'. At still greater distances, beyond the limits of the 100 000 light-year disc, astronomers have discovered other huge collections of stars, gas, dust and dark matter that, like the Milky Way, occupy relatively well-defined, and usually well-separated, volumes of space. These structures are called **galaxies**. Figure 1.2 shows one of these 'external' galaxies that is thought to be somewhat similar to the Milky Way in many respects. The Milky Way is simply *our* galaxy, the galaxy in which we have been born and have come of age. To emphasize this, the Milky Way is often referred to as 'the Galaxy', the capital 'G' distinguishing it from the billions of other galaxies in the observable Universe. In later chapters you will learn more about those other galaxies. Here in Chapter 1 we concentrate on the Milky Way.

Figure 1.2 An 'external' galaxy (NGC 2997) thought to be similar to the Milky Way. If this really represented the Milky Way, the Sun would be located about halfway between the centre and the edge of the flattened 'disc' of stars, gas and dust. This location, within a relatively thin disc, explains why we mainly see the Milky Way as a 'band' encircling the Earth. (D. Malin/AAO)

In the sections that follow you will learn a great deal about the Milky Way. Section 1.2 provides a general overview of the Milky Way as a galaxy, including its structure, size and composition. Section 1.3 is devoted to the mass of the Milky Way. Sections 1.4 and 1.5 discuss some of the main structural components of the Galaxy in detail, and Section 1.6 considers the formation and evolution of the Milky Way. As you study these sections you will also gain insight into the *process* of astronomical science. You will see, in outline at least, *how* the nature of the Galaxy has been uncovered, *how* the disc of the Milky Way has been shown to be about 100 000 light-years across, *how* the Milky Way's mass has been roughly determined to be about $10^{12} M_{\odot}$ and *how* astronomers have estimated the age of the Galaxy at between 12×10^9 and 15×10^9 yr. By examining the process of astronomy you will begin to understand how the making of careful observations, combined with the continuous review of their significance in the light of physical laws, enables findings and theories to be critically examined, refined and improved. Developing an understanding of this process is more important than learning any particular fact or figure.

1.2 An overview of the Milky Way

This section examines some of the general features of the Milky Way as a galaxy. It starts with an introduction to the main structural components of the Milky Way, and then goes on to examine the sizes of those components and their constituents. Particular emphasis is given to the stellar content of the Milky Way, since it is the study of the nature, distribution and movement of the stars in the Milky Way that

has provided the key to much of what has been learnt about the Galaxy. The discussion of stars in the Milky Way introduces the important concept of *stellar populations*, and the section ends with a brief discussion of the related issue of the chemical evolution of the Milky Way. Detailed discussion of the mass of the Milky Way is deferred to Section 1.3.

1.2.1 The structure of the Milky Way

When asked to describe the Galaxy, most people think of the huge numbers of stars gathered together to form a flattened disc. Indeed, this was the description we began with above. There are some 10^{11} stars in our Galaxy, most of them in the disc, and they dominate the visible light emitted from it, so this view is quite understandable. However, while stars are luminous and hence easily seen, they are only the visible tips of our Galactic iceberg. There is more to the Galaxy, *vastly* more, than meets the eye.

The modern view of the Galaxy is that its largest and most massive component is a huge, roughly spherical cloud consisting of some kind of non-luminous matter. Since this non-luminous matter has never been directly observed at any wavelength it is known as **dark matter**. The nature of dark matter is unknown at the time of writing. Its presence is revealed by the gravitational influence that it has on the more familiar forms of matter that can be directly observed. Despite knowing very little about dark matter, most astronomers have become convinced that the total mass of dark matter in the Milky Way is about ten times greater than the total mass of stars and about 100 times greater than the total mass of gas and dust. Furthermore, this appears to be the case for other galaxies too; as you will see later, the nature, distribution and significance of dark matter is a recurring theme of great importance in the study of galaxies and cosmology.

The huge cloud of dark matter that is believed to be the main structural component of the Milky Way is usually referred to as the **dark-matter halo**. The mass of this component is so great that it is the gravity of the dark matter, rather than the gravity of all of the stars, that is primarily responsible for holding the Galaxy together. The gravitational influence of dark matter on luminous material allows astronomers to infer the shape of the dark-matter halo. They have concluded that the dark-matter halo takes the form of a *spheroid* – the three-dimensional figure formed by rotating an ellipse about one of its axes. Spheroids themselves come in a variety of shapes; the dark-matter halo is thought to be an *oblate* spheroid, that is to say it resembles a sphere that has been flattened at its poles, with a shortest-to-longest axis ratio of about 0.8. The stars of the Milky Way, which provide most of the Galaxy's luminous output are located in the centre of this dark-matter halo.

Let's now consider in more detail how this luminous matter is distributed – and how it makes up the visible parts of the Galaxy. Most of the stars, including the Sun, occupy a flattened, disc-shaped volume. This is called the **Galactic disc**, or simply the **disc**, and is the second major structural component of the Galaxy. Its mass is about $10^{11}M_{\odot}$. In addition to stars, the Galactic disc contains a substantial mass of gas and dust in the space between the stars. The mid-plane of the disc defines the **Galactic plane**, which plays an important part in providing a system of coordinates for defining positions in the Milky Way (see Box 1.1). The Sun is located very close to the Galactic plane, about halfway between the centre of the disc and its outer edge.

BOX 1.1 GALACTIC COORDINATES

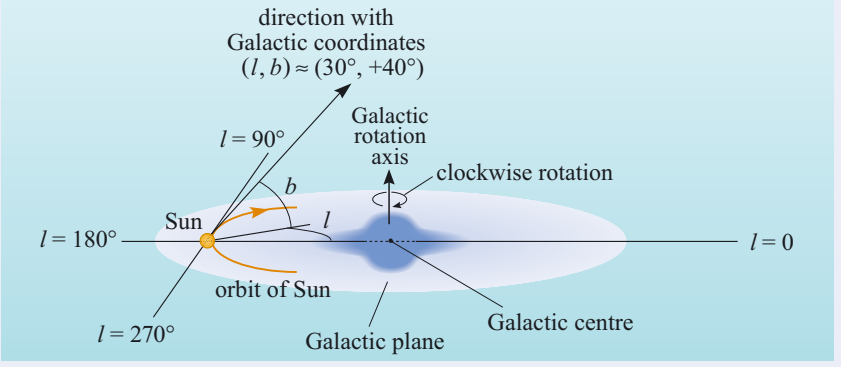


Figure 1.3 Galactic coordinates are centred on the Sun, and use the Galactic plane and the direction approximately towards the Galactic centre to define a frame of reference. A direction with $l \approx 30°$ and $b \approx +40°$ is shown.

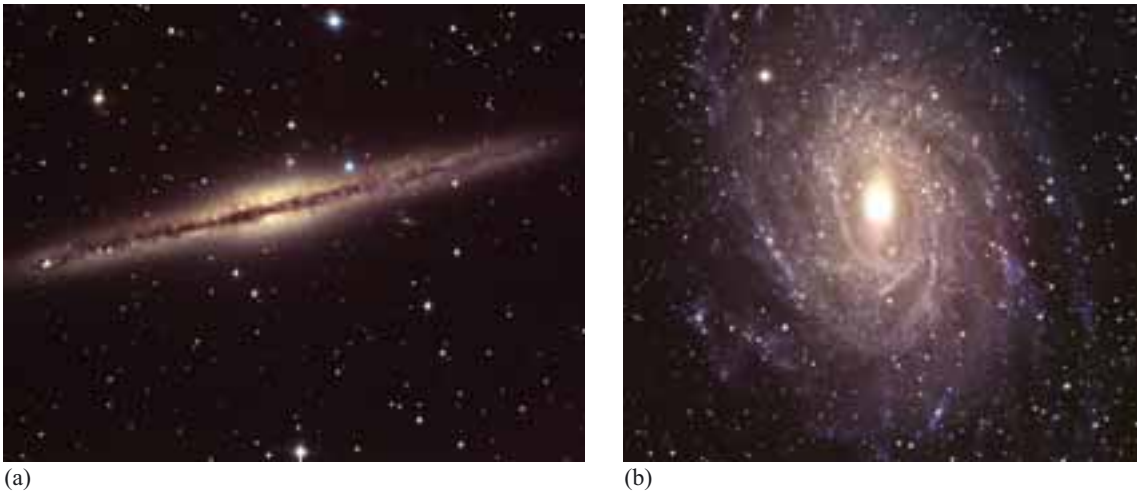
Astronomers who study the Galaxy find it useful to define a coordinate system that reflects the Galaxy's symmetry about the Galactic plane as it is viewed from the Earth. In the system of **Galactic coordinates**, the direction of any object in the sky can be expressed in terms of its **Galactic latitude** (b) and **Galactic longitude** (l), both of which are angles normally expressed in degrees. Figure 1.3 shows how they are defined. The **Galactic equator** runs close to the mid-plane of the Milky Way's disc. The origin of the coordinate system, the point $l = 0°$ and $b = 0°$, is defined to be *nearly* – but not precisely – in the direction of the Galactic centre, which is in the direction of the constellation Sagittarius. (The convention is to show the Galaxy with the North Galactic Pole upwards.) Galactic latitudes are measured north (b positive) or south (b negative) of the Galactic equator, so they range between $b = +90°$ and $b = -90°$. Galactic longitude ranges from $l = 0°$ (roughly towards the Galactic centre), eastwards through $l = 90°$ (roughly in the direction of Galactic rotation), and on to $l = 360°$.

From our vantage point, we see the disc edge-on from within. It is our location that causes most of the other stars in the Milky Way's disc to appear concentrated within a band looping around us – the source of the diffuse band of light that Galileo resolved into stars.

If the disc could be viewed face-on, we would see a spiral pattern due to the presence of bright features called **spiral arms**. (Such arms are an obvious feature of many galaxies; the evidence for their existence in the Milky Way is described in Section 1.4.) Edge-on and almost face-on views of two galaxies that are thought to be broadly similar to the Milky Way are shown in Figure 1.4.

The overall spiral shape in Figure 1.4b is clear, but in detail the arms are fragmented and distorted. Although spiral arms are prominent, they stand out because they contain unusually hot, *luminous* stars, and not because they contain unduly large numbers of stars. The reasons why bright stars are concentrated in this way is discussed in detail in Section 1.4. Spiral arms suggest that galaxies are rotating. Sure enough, the Sun and its neighbouring stars in the disc orbit the Galactic centre at speeds of about 220 km s^{-1}. However, as you will see later in this chapter, the stars and the pattern of spiral arms generally travel at *different* speeds.

Figure 1.5 is a schematic diagram of the major structural components of the Galaxy that we are introducing in this section, including the dark-matter halo and disc that we have met already, as well as other components that are described shortly. Their approximate sizes, which we discuss in Section 1.2.2, are also given. This diagram should be compared with the edge-on and face-on images of galaxies in Figure 1.4, which are broadly similar to the Milky Way.



(a) (b)

Figure 1.4 (a) NGC 891, a spiral galaxy seen edge-on, observed in infrared light. (b) NGC 6744, a barred spiral galaxy seen almost face-on. These galaxies are thought to be similar in structure to the Milky Way. ((a) J. C. Barentine and G. A. Esquerdo, Kitt Peak, NOAO; (b) S. Lee, C. Tinney and D. Malin/AAO)

Towards the centre of the Galaxy the density of stars increases, and the Milky Way appears to be 'thicker' there than further out. This broad, central region is called the **bulge**, and is the third structural component of the Galaxy. An example of a bulge can be seen in the centre of the galaxy pictured in Figure 1.4a. The Milky Way's bulge has a mass of around $10^{10} M_{\odot}$. The central regions of the Galaxy are notoriously difficult to observe from our vantage point near the Sun, but there is evidence that the Milky Way's bulge is elongated, which makes the Milky Way a **barred spiral galaxy** like the one in Figure 1.4b. The Milky Way is far from being unique in this respect; observations of other spiral galaxies show that most have *some* trace of a central bar.

Surrounding the disc is a sparsely populated structural component called the **stellar halo**, or just the **halo**, the mass of which is only about $10^{9} M_{\odot}$. Because the number of stars per unit volume is much lower in the halo than in the disc, the halo does not show up in the two images in Figure 1.4. Each clearly shows a disc and a bulge, but gives little indication of a stellar halo, which is only revealed by more detailed studies. This Galactic component is not flat like the disc, but rather has a spheroidal shape, with only slight flattening. For this reason, the halo (and often the bulge with it) is sometimes referred to as the **Galactic spheroid**, or simply the **spheroid**. The name 'stellar halo' suggests there are only stars, not gas, in this component. In fact there is some gas, but present only at a very low density. Furthermore, as we see in Section 1.6, some of the gas traditionally associated with the halo may not really belong to the Galaxy at all. This highlights the fact that the Galaxy is not as isolated in space as the introduction above may have suggested; rather the Galaxy is surrounded by an environment with which it interacts. More will be said about this *intergalactic medium* in later chapters, but we shall also see several other examples of the way the Galaxy interacts with its environment later in this chapter.

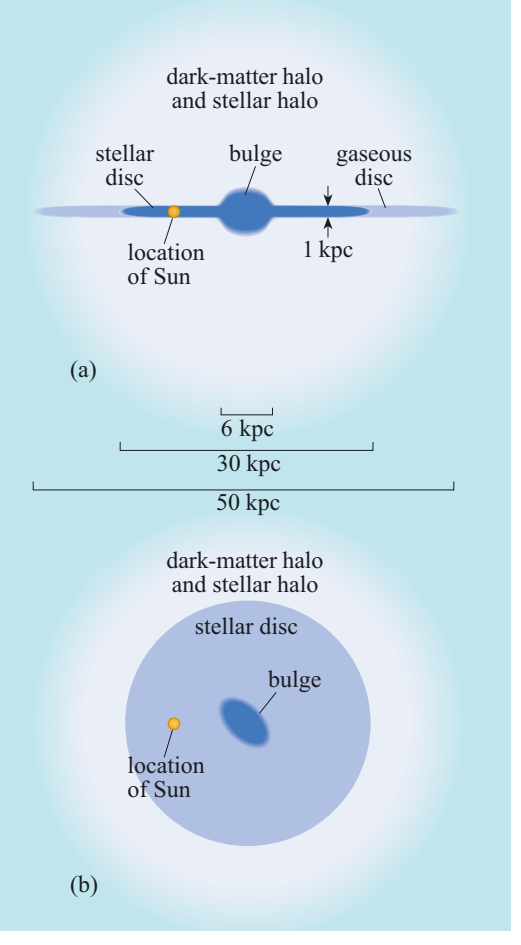


Figure 1.5 (a) Edge-on and (b) face-on schematic views of the four major structural components of the Milky Way: the dark-matter halo, the disc, the stellar halo and the bulge. The sizes indicated in this figure are expressed in kiloparsec (kpc), where 1 kpc ≈ 3260 ly.

In summary, the Galaxy's major component is the *dark-matter halo*. Embedded within this is the *Galactic disc*, which is where most of the stars, gas and dust are found. The central region of the Galaxy is thicker than the rest of the disc, and is called the *bulge*. Surrounding the disc is the sparsely populated *stellar halo*. The disc contains bright *spiral arms*, and the bulge is elongated into a *bar*; the Milky Way is therefore a *barred spiral galaxy*.

■ What are the shapes and approximate masses of each of the four main structural components of the Galaxy?

❑ The dark-matter halo and the stellar halo are both slightly flattened (oblate) spheroids. The disc has the flattened circular form its name implies, and the central bulge is elongated into a bar. Very roughly, the stellar halo has a mass of $10^9 M_\odot$, the bulge mass is $10^{10} M_\odot$, the mass of the disc is $10^{11} M_\odot$ and the dark-matter halo has a mass of $10^{12} M_\odot$, although this last value is particularly uncertain.

1.2.2 The size of the Milky Way

A table of frequently used conversion factors and physical constants is provided in the Appendix.

Having introduced the main structural components of the Galaxy, we now examine their sizes. Although the size of the disc was given above as 100 000 light-years, distances in the Galaxy are usually measured in units of **parsecs** (pc) or **kiloparsecs** (kpc), where 1 kpc = 1000 pc. One parsec is equal to about 3.26 light-years, or 3.09×10^{16} m.

So, how big is the Galaxy? The answer depends greatly on which component you measure. The dark-matter halo is the most extensive component, but it is also the most difficult to measure since its presence is deduced only from its gravitational influence. The size of the dark-matter halo can be assessed from its effect on the motions of neighbouring galaxies. The Magellanic Clouds are two small galaxies that are 50 to 60 kpc from the Milky Way, and the dark-matter halo apparently extends at least that far. So, in answer to the question: 'How big is the Galaxy?' you could cite the distance of the Magellanic Clouds as a lower limit on the radius of the dark-matter halo, implying a diameter of at least 100 to 120 kpc.

Since the dark-matter halo cannot be observed very easily, you may prefer to consider a different question: 'How big is the disc of the Galaxy?' It turns out that even this more carefully posed question requires a cautious response. In short, the answer depends on which constituent of the disc you measure: stars or gas. The stellar disc of the Milky Way has a radius of at least 15 kpc. Observations indicate that the Sun is about 8.5 kpc from the centre of the Milky Way, which places it around halfway out in the stellar disc. This disc is around 1 kpc thick, which means some stars travel up to about 500 pc from the mid-plane of the disc.

■ If the radius of the Galactic disc is 15 kpc, then its diameter is 30 kpc. How many light-years is 30 kpc?

❑ 1 pc = 3.26 ly, so 30 kpc = 30×10^3 pc $\times$ 3.26 ly pc^{-1} = 9.8×10^4 ly $\approx$ 100 000 ly.

In contrast to the stars, the *gas* and in particular the atomic hydrogen in the Galactic disc extends out to a radius of at least 25 kpc (although its density does fall considerably beyond 15 kpc). A clear example of the difference between the gaseous

and stellar discs in a spiral galaxy is given by the barred spiral galaxy NGC 6744, which we observe almost face on, and which we believe is similar to the Milky Way. This was the galaxy pictured in Figure 1.4b. Contours showing the atomic hydrogen gas density inferred from radio measurements of NGC 6744 are superimposed on the optical image in Figure 1.6. It has a gaseous disc that extends out to at least 1.5 times the radius of its stellar disc, maintaining the spiral structure seen in the visible image as it does so.

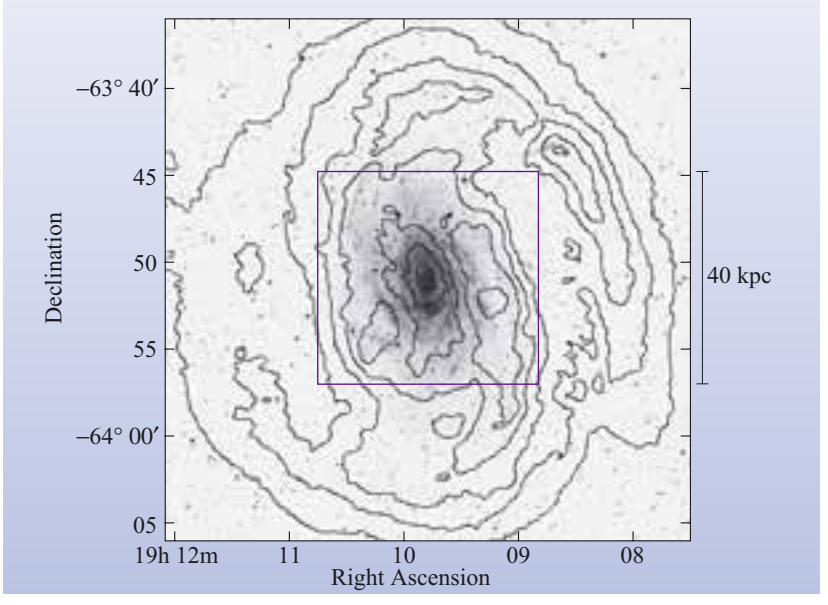


Figure 1.6 Contours of atomic hydrogen gas density in the barred spiral galaxy NGC 6744, based on radio observations, superimposed on an optical image from the Digitized Sky Survey. The central square corresponds to the image in Figure 1.4b, and measures approximately 40 kpc on each side. Note that the gas extends well beyond the stellar disc, and that the spiral pattern is visible out to the edge of the gas disc. This galaxy is believed to be similar to the Milky Way. (Ryder *et al.*, 1999)

■ Can you suggest why the stars in the Galactic disc are found only out to a radius of 15 kpc from the Galactic centre, even though the gaseous disc extends out to 25 kpc?

❑ Stars form from the gas, and they form preferentially where the gas density is higher. Even though the gas extends out to 25 kpc, its density falls considerably beyond 15 kpc. So a plausible explanation may be that star formation does not occur at the gas densities that occur at a radius greater than about 15 kpc.

As the stellar halo has no substantial gaseous component, its size is given by the distribution of stars. However, defining the edge of this distribution is difficult, because the density of stars falls off gradually with distance from the centre of the Galaxy. For now we simply state that the stellar halo extends further than the disc, well beyond 20 kpc.

The bulge of the Galaxy takes the form of an elongated bar. The longest axis of this bar is in the Galactic plane and stretches out to about 3 kpc either side of the Galactic centre. The cross-sectional diameter of the bar is roughly 2 kpc. The mass of the bulge is much greater than that of the stellar halo, but its small size means that it has little relevance to any discussion of the overall size of the Galaxy.

Although approximate sizes for each of the Galaxy's main structural components have now been quoted, you may have noticed that the exact size of our Galaxy has still not been specified. It is always possible to define the size of the Galaxy as the size of *one* of the components, but such a definition would be rather arbitrary. It is more useful to retain a broad knowledge of the nature and scale of each of the structural components that make up our Galaxy.

QUESTION 1.1

If the Sun is 8.5 kpc from the Galactic centre and moving in a circular orbit at 220 km s^{-1}, how long will it take to travel once around the Galaxy? Express your answer in both SI units (seconds) and years.

(Recall that the relationship between a body's speed, v, the distance travelled, d, and the time taken, t, is $v = d/t$, and that the circumference of a circle of radius R is $d = 2\pi R$.)

1.2.3 The major constituents of the Milky Way

The major structural components of the Milky Way – the dark-matter halo, disc, bulge and stellar halo – have now been introduced, and you have briefly encountered their constituents: dark matter, stars, gas, and dust. Now we look at these constituents more closely. We start with the dominant component, the dark matter.

Dark matter

Dark matter is detectable by its gravitational influence, but appears neither to emit nor absorb light nor any other form of electromagnetic radiation. The total mass of dark matter in the Milky Way seems to be about $10^{12} M_{\odot}$. However, at present we do not know the nature of this matter.

Everyday matter that we regularly encounter on Earth, and which constitutes the material in ordinary stars, is primarily made up of particles called **baryons**, the best known of which are protons and neutrons. Some of the dark matter may be dense, non-luminous, baryonic matter, but there is strong evidence (related to studies of the early Universe that are described later) indicating that much of the dark matter is non-baryonic. So, in addition to **baryonic dark matter**, which would be 'ordinary' matter that happened to be difficult to detect and about which we know very little, there is also a need for **non-baryonic dark matter**, about which we know even less.

Although the nature of non-baryonic dark matter is still a mystery, a number of proposals have been made regarding its possible composition. Most of these proposals assume that the non-baryonic dark matter consists of hypothetical fundamental particles of one kind or another. If any of these proposals is correct, then the dark-matter halo would simply be a vast cloud of these particles. Furthermore, since we are situated within this cloud, dark-matter particles should be detectable here on Earth. As we shall see in Chapter 8, experiments with this aim are underway or currently being developed.

Stars

There are about 10^{11} stars in the Galaxy, and, since the Sun's mass ($M_{\odot} \approx 2 \times 10^{30}$ kg) is typical, they have a combined mass of about $10^{11} M_{\odot}$, roughly one-tenth the total mass of dark matter. The vast majority of these stars occupy the disc.

Stars can differ from one another in their *mass*, their *age*, and their *chemical composition*; differences in these three fundamental parameters lead to differences in other properties, such as luminosity and temperature. The *spectral class* of a star is an important property that is closely related to its temperature. In order of

decreasing surface temperature, the spectral classes are O, B, A, F, G, K, M. Many of the brightest stars in the Milky Way are large, bright, blue–white stars belonging to classes O and B, but by far the most common are small, faint, red stars belonging to class M. Stars form from large clouds of gas, and spend the greater part of their luminous lives as *main sequence stars* that are powered by the conversion of hydrogen into helium in their cores. The subsequent evolution of a star depends on its mass (the greater the mass, the shorter the life), but in most cases it includes the eventual enlargement of the star to form a *giant*, and the conversion of helium into a range of heavier elements.

Do the stars in each of the three Galactic components that contain stars – the stellar halo, disc and bulge – have the same range of masses, ages and compositions? Observations show that they do not. It turns out that the stellar halo and bulge contain much older stars than the disc, and that stars in the halo contain few elements heavier than helium. These differences led to categories of stars called **stellar populations** being defined. The differences between the populations tell us much of what we know about the origin and evolution of the Milky Way. We examine the differences between the various stellar populations in Section 1.2.4, and we shall return to their evolutionary implications several times in this chapter.

Gas and dust

Most of the Milky Way's gas and dust lies in the disc, and is found within a vertical distance of 150 pc of the Galactic plane: it does not extend nearly so far from the mid-plane of the Galaxy as do the stars.

The gas is roughly 70% hydrogen and 28% helium (by mass). The remaining 2% is made up of the other elements – these are collectively referred to (by astronomers) as **metals**. The hydrogen can exist in various forms depending on the density, temperature, and flux of ultraviolet (UV) radiation in each locality. In high-density, low-temperature environments with a low UV flux, the hydrogen is mostly in the form of **molecular hydrogen** (H_2). In environments where the temperature and/or the UV flux is high enough to free the hydrogen atom's single electron, there is a likelihood of finding **ionized hydrogen** (H^+, usually written HII and pronounced 'H-two'), particularly where the density is low enough to reduce the chance of the liberated electrons recombining with the positive ions. **Atomic hydrogen** (H, often written HI and pronounced 'H-one') occurs where conditions lie between the other two extremes. The total mass of gas in the Galaxy is estimated to be about 10% of the stellar mass.

What astronomers call **dust** consists of tiny lumps of solid (condensed) compounds of carbon, oxygen, silicon and other metals. (The term 'metals' is used here, and throughout this book, in its astronomical sense.) Most of the bulk of a dust grain comprises either graphite or silicate compounds. The outside of the grain is often surrounded by a coating, or *mantle*, of more volatile compounds such as water-ice (H_2O), ammonia (NH_3) and carbon monoxide (CO). Dust particles are typically 10^{-7} to 10^{-6} m (0.1 μm to 1 μm) across, close in size to smoke particles on Earth. The total mass of dust in the Galaxy is about 0.1% of the stellar mass.

The size of dust particles is comparable to the wavelength of light, and thus makes them particularly effective at scattering light, as well as absorbing it. Since the dust is mainly concentrated within 150 pc of the mid-plane of the disc, its obscuring effect is especially evident when we look in directions close to the

Galactic plane, as indicated in Figures 1.1 and 1.4a. The presence of dust in the disc of the Milky Way severely limits our ability to make optical observations in certain directions (see Figure 1.7), and creates what is called the **zone of obscuration** or **zone of avoidance** on the sky – a band, extending to about 15° either side of the Galactic equator, within which very few galaxies are seen. Fortunately, technological developments over the last few decades have allowed astronomers to make observations at infrared and other wavelengths that are relatively unaffected by dust obscuration.

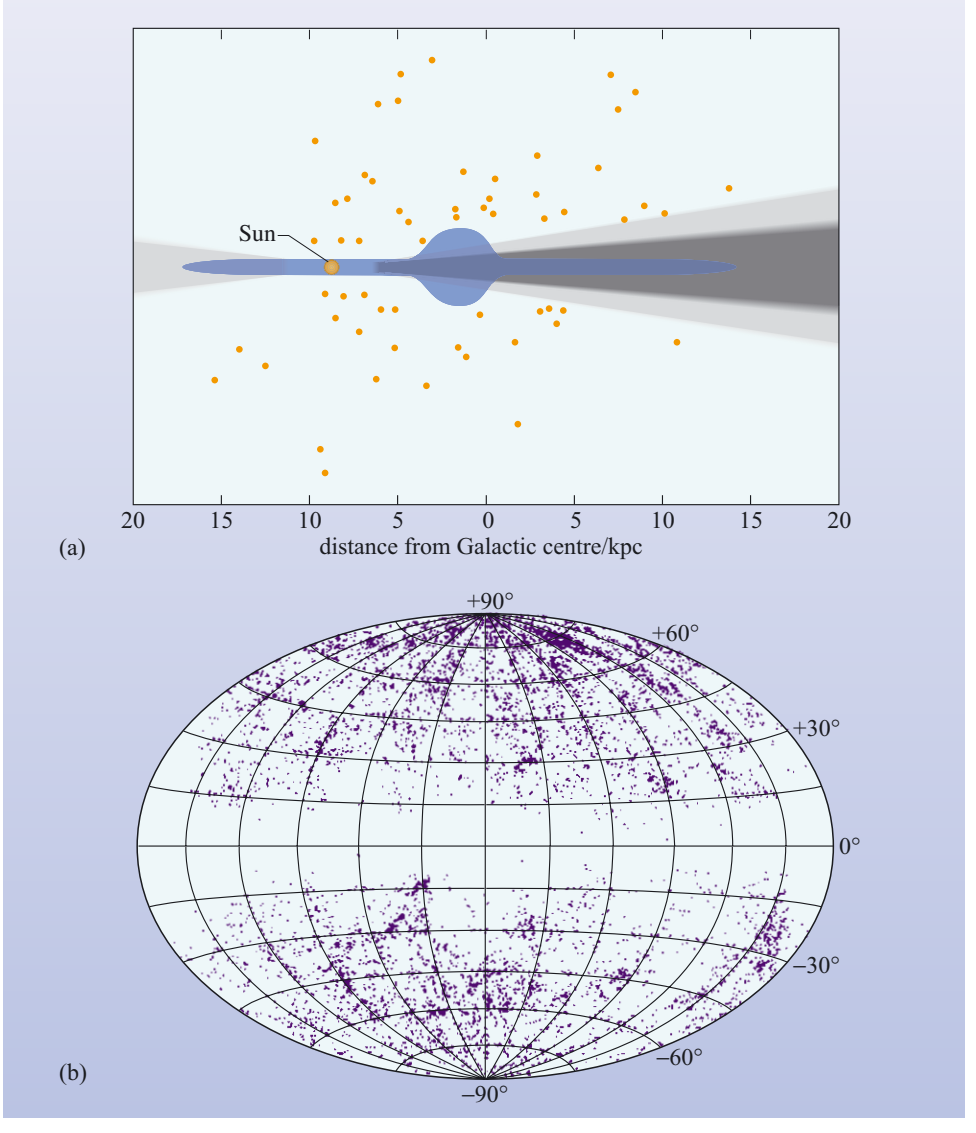


Figure 1.7 (a) Due to the presence of dust in the Milky Way's disc, optical observations in the regions shaded light grey are very difficult, and those in the dark grey region are essentially impossible, except for *very* bright objects within about 5 kpc. (b) An all-sky map in Galactic coordinates, in which the Galactic equator runs horizontally across the middle and the centre of the map corresponds to the direction directly *away* from the centre of the Galaxy ($l \approx 180°$, $b \approx 0°$). This map shows the positions of bright galaxies beyond the Milky Way. Dust in our Galaxy prevents us from seeing the light from other galaxies in directions close to the Galactic plane. This illustrates the effect of the zone of avoidance. ((b) Binney and Merrifield, 1998)

■ What is the total mass of gas in the Galaxy, in solar masses?

❑ The mass of gas is 10% of the stellar mass of the Galaxy, and the latter is about $10^{11}M_\odot$, so the mass of gas is

$$10\% \times 10^{11}M_\odot = (10/100) \times 10^{11}M_\odot = 10^{-1} \times 10^{11}M_\odot = 10^{10}M_\odot$$

■ What is the total mass of dust in the Galaxy, in solar masses?

❑ The mass of dust is 0.1% of the stellar mass of the Galaxy, and the latter is $10^{11}M_\odot$, so the mass of dust is

$$0.1\% \times 10^{11}M_\odot = (0.1/100) \times 10^{11}M_\odot = 10^{-3} \times 10^{11}M_\odot = 10^{8}M_\odot$$

The term **interstellar medium**, or **ISM**, is used to describe the gas and dust that occupies the space between the stars. On average the ISM contains about 10^6 particles per cubic metre, but the density and nature of the ISM varies greatly from one region to another, so this average does not have major significance. Almost half of the ISM (by mass) is contained in cool **dense clouds**, often called **molecular clouds** because they are rich in molecular hydrogen (H_2). These clouds occur with a wide range of masses, the most massive containing up to $10^7M_\odot$ of gas and dust. Molecular clouds are usually many thousands of times denser than the average ISM, with typical diameters of 10 to 100 pc. However, these clouds account for only 1% or so of the volume of the ISM, despite contributing nearly 50% of its mass. Another contribution to the ISM comes from localized clouds (typically a few parsecs across) of hot, ionized gas known as **HII regions**. These regions account for a few per cent or so of the ISM's mass and volume. The Orion Nebula (see Figure 1.8) is one of the best known of these regions.

Figure 1.8 The Orion Nebula (M42). A prominent HII region in which the hydrogen is ionized by the UV radiation from a group of bright young stars. The Orion Nebula is about 6 pc in diameter and contains hundreds of solar masses of gas. The nebula forms a sort of blister on the face of a giant molecular cloud. (NASA)

Much of the remaining volume of the ISM is occupied by an **intercloud medium** that may be hot or warm, depending on local conditions. Within the Galactic disc, the intercloud medium can be loosely regarded as a disc about 300 pc thick that is rich in atomic hydrogen (HI), and more or less uniformly distributed. However, this disc of intercloud medium, along with the thousands of individual clouds it contains, is embedded in a large body of hot intercloud medium that occupies at least part of the halo. There is still much uncertainty about the distribution of the intercloud medium, but the low density of this component of the ISM means that it contributes only a minor part of the ISM's mass, despite its large volume.

The ISM is intimately associated with star formation and stellar evolution: stars are born from cool dense molecular clouds, and they return matter to interstellar space in a variety of ways that gradually increase the proportion of heavy elements in the Galaxy. This process is known as *Galactic chemical enrichment*, and is discussed in greater detail in Section 1.2.5.

We have now completed our survey of the main Galactic components and their constituents. Having seen that the Galaxy is made of dark matter, stars, gas and dust, and how it is divided into the dark-matter halo, disc, stellar halo and bulge, we are in a position to start using observations of these entities to uncover the history of the Galaxy. We begin this process by looking more closely at the differences between populations of stars.

1.2.4 The stellar populations of the Milky Way

It was noted above that a star can be described by its mass, age and composition, and we foreshadowed the fact that significant differences in these parameters exist between the stars in each of the different Galactic components. In this section we explore the nature of those differences and what they tell us about the evolution of the Galaxy.

The first indication that there might be systematic variations in stellar properties between one region of the Milky Way and another came from a comparison of observations of our Galaxy with the nearby spiral galaxy M31 (the Andromeda Galaxy). However, before we attempt to describe and understand these differences it is necessary to know something about the spherical clusters of stars called **globular clusters** (see Figure 1.9). Globular clusters are compact, dense clusters of very old stars, typically containing 10^4 to 10^6 members in a spherical region of space less than about 50 pc in diameter. Around two-thirds of globular clusters belong to the stellar halo and one-third to the disc. While globular clusters are easy to recognize, they account for only about 1% of all stars in the stellar halo. The distribution of globular clusters in the Milky Way is shown in Figure 1.10, which gives a good indication of the shape of the stellar halo and how fundamentally it differs from the disc.

Figure 1.9 The globular cluster 47 Tuc. This globular cluster is one of the 150 or so globular clusters that are known to be associated with the Milky Way. (NASA/ESA)

The German–American astronomer Walter Baade (Figure 1.11) found that he could just detect the brightest O and B stars and red giants in M31 as single stars in his photographic observations, despite that galaxy being 750 kpc away. He noted a difference in colour between the stars in the disc and spheroid of M31. The disc stars were blue, while the spheroid stars were red. He named these two stellar types **Population I** (Pop. I, pronounced 'pop one') and **Population II** (Pop. II, pronounced 'pop two'), respectively. Importantly, he noted that the colour difference meant that Pop. I stars resembled the brightest stars in the disc of the Milky Way, which are predominantly blue, while the Pop. II stars resembled the brightest stars in the Milky Way's globular clusters, which are predominantly red.

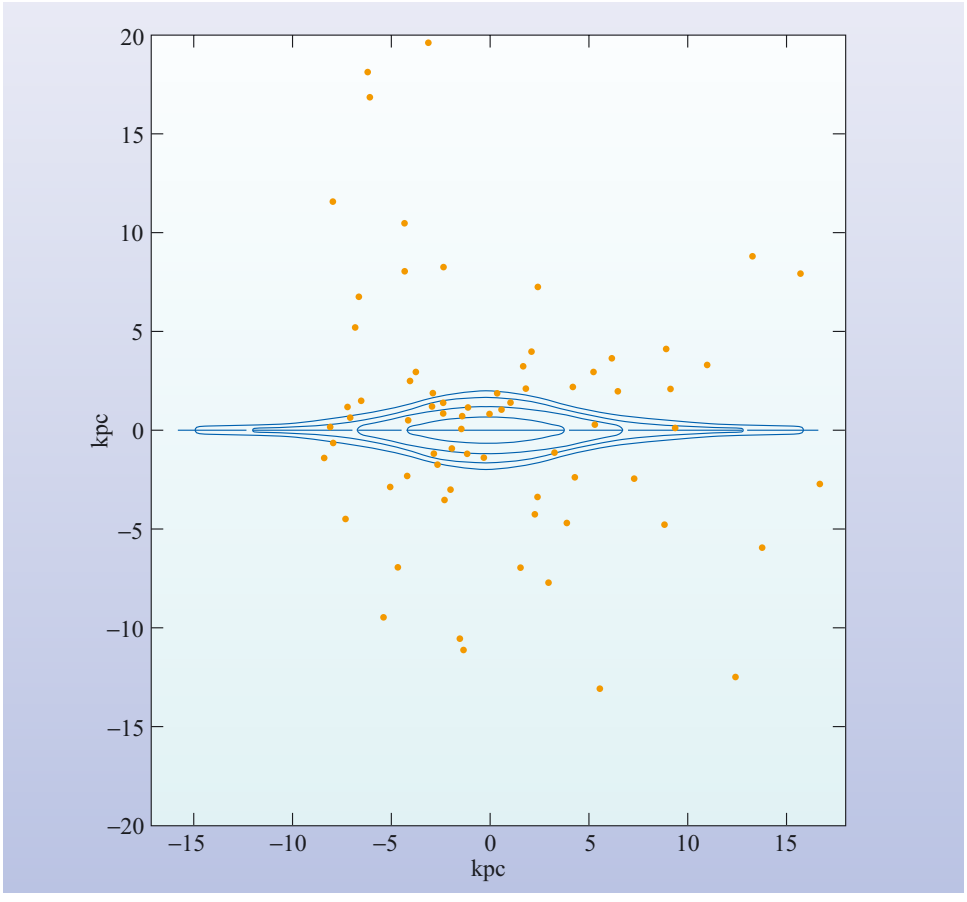


Figure 1.10 An edge-on projection of the Galaxy showing the locations of a sample of globular clusters. They are distributed approximately spherically about the centre of the Galaxy, in contrast to the highly flattened shape of the disc. The contours show curves of equal density of disc stars. The Sun is located 8.5 kpc from the Galactic centre (at the distance −8.5 kpc from the Galactic centre in this diagram). (Blaauw and Schmidt, 1965)

WALTER BAADE (1893–1960)

Walter Baade (Figure 1.11) was educated in Germany, at the universities of Münster and Göttingen. In 1931, after spending a number of years at the University of Hamburg, he moved to the USA where he worked successively at the Mount Wilson and Mount Palomar Observatories in California. It was during this period that Baade carried out the astronomical work for which he is best remembered. This includes his recognition of the existence of two distinct stellar populations, and the discovery that many of the measured distances to other galaxies were incorrect because two different populations of Cepheid variable (see Chapter 2) had been treated as a single uniform population. Baade also discovered a one-sided jet in the galaxy M87, and recognized the Crab Nebula as the remnant of a supernova that had been observed in 1054. Baade later returned to Germany to take up a post at Göttingen, and died there in 1960.

Figure 1.11 Walter Baade. (Mt. Wilson Observatory)

Following on from Baade's discovery, much effort has gone into studying the characteristics of the different populations. Advances in observational technology and astrophysical theory have made it possible to refine Baade's original idea, so that stellar populations can now be defined in terms of the age, metal content, and location of the stars. While this work has been extremely profitable in developing an understanding of the nature of the stars and the origins of the populations, the proliferation of alternative definitions of populations has not always been helpful. As a result, population descriptions are sometimes used loosely, often meaning different things to different astronomers. This problem has been exacerbated by correlations that often exist between the three key population parameters: age, metal content, and location. In this chapter we adopt one of the simpler classification schemes, concentrating on the two major population divisions that originated with Baade. We also note a third population, unknown to Baade and which has not yet been observed, but which almost certainly existed early in the history of the Universe.

- The three key population parameters in current use are age, metal content, and location, but Baade observed colours and locations. Which of these key parameters corresponds to colour, and why?
- The colours of very luminous stars depend on their age: luminous blue stars are short lived so they are all young; luminous red stars are old, they are elderly red giants.

Before defining the three populations that are used in this book it is helpful to define a quantitative measure of a star's metal content. A useful definition, although by no means the only one possible, provides a measure of the fraction of the mass of an object that is accounted for by elements heavier than helium, that is, the metals. This is the **metallicity**, Z, defined by the following equation.

$$Z = \frac{\text{the mass of elements heavier than helium in the object}}{\text{the mass of all elements in the object}} \qquad (1.1)$$

The metallicity of the Sun is $Z = 0.02$. That is, 2% of the Sun's mass comes from elements heavier than helium.

Having set out the key parameters for describing stellar populations, we now go on to see how these differ for the three populations of interest.

Population I

Pop. I, which Baade associated with the disc, includes many very young stars, some just a few million years old, but also includes some as old as 10^{10} yr. Their metallicities are mostly in the range $Z = 0.01$ to 0.04 (i.e. within a factor of two of the solar value) although some Pop. I stars with still lower metallicities exist. The stars of Pop. I move in essentially circular orbits (around the Galactic centre), which do not take them far above or below the plane of the Galaxy, and hence they are confined to the flat, circular structure that constitutes the Galactic disc. The answer to Question 1.1 shows that the Sun takes roughly 240 million years to move once around the Galaxy. This value is typical of other Pop. I stars near the Sun.

Population II

Pop. II stars occupy the spheroid – the stellar halo and bulge – and turn out to be the oldest stars known, with ages in the range (12 to 15) $\times$ 10^9 yr. Conspicuous examples are globular-cluster stars. Little or no interstellar gas is still associated with Pop. II stars, which is consistent with star formation in the spheroid ceasing long ago. Because this population is so old, only low-mass stars (which have long lifetimes) still shine as main sequence stars burning hydrogen in their cores. The more massive stars that formed at the same time as the surviving low-mass ones have already left the main sequence and are now red giants or white dwarfs.

For a long time it was thought that all Pop. II stars had much lower metallicities than do Pop. I stars, but it is now known that this applies only to the stellar halo, for which $Z < 0.002$, and where the lowest-metallicity stars detected so far have $Z \sim 2 \times 10^{-6}$. Some bulge stars, on the other hand, have the same metallicity as the Sun.

Unlike disc stars, Pop. II stars do not follow circular orbits, nor are they confined to the plane of the Galaxy. They move in eccentric orbits (see Figure 1.12), although still attracted to the Galactic centre, and may travel many kiloparsecs from the Galactic plane. This is of course consistent with Pop. II stars belonging to the spheroid, briefly passing through the disc as they move from one side of the Galactic plane to the other. Such fleeting visitors to the Galactic disc are known as **high-velocity stars** because of their high speed relative to the Pop. I stars that belong to the disc. In contrast to the disc, there is almost no net rotation of the halo, so almost half of all halo stars travel in **retrograde** orbits (i.e. in the opposite sense to the more orderly disc stars, which all orbit in the clockwise direction as viewed from the north Galactic pole).

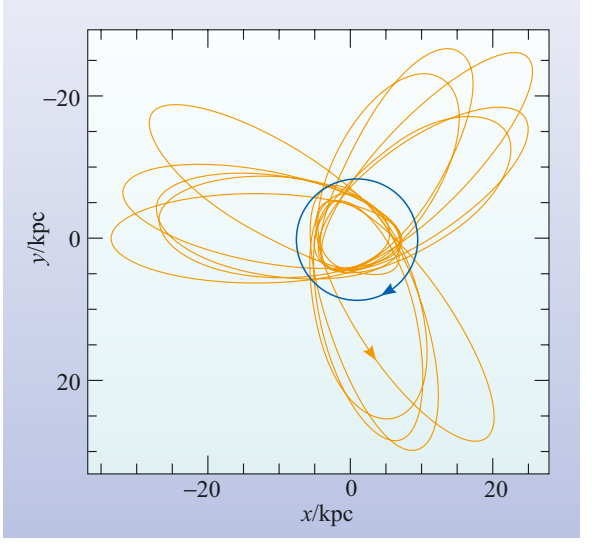


Figure 1.12 A face-on view of the Galaxy showing sample orbits for Pop. I (blue) and Pop. II (orange) stars. Shown is a clockwise, circular orbit for a Pop. I star 8.5 kpc from the Galactic centre, and an anticlockwise (retrograde), three-lobed orbit for a Pop. II star that takes it from 4 kpc to 35 kpc from the centre. This Pop. II star also travels up to 20 kpc above and below the disc, while the Pop. I star remains close to the Galactic plane. (S. Ryan (Open University))

Population III

The term Pop. III describes a theoretical population rather than one that has actually been observed. It encompasses stars that formed out of the unprocessed gas that would have been produced by the big bang. This material would have been almost entirely composed of hydrogen and helium. Even lithium, the next most abundant element in this gas, would only constitute one particle in a billion. Consequently the metallicity of these stars when they first formed ($Z \approx 10^{-9}$, see Example 1.1), would have been much lower than that of even the lowest metallicity stars of Population II ($Z \approx 2 \times 10^{-6}$). No Pop. III star has yet been observed, but theoretically they must have existed as the very first generation of stars in the Universe.

EXAMPLE 1.1

The lithium nuclei that were produced in the big bang account for only a tiny fraction of the total number of nuclei produced, around 1.6×10^{-10} of the total number. Given that each lithium nucleus has a mass seven times that of hydrogen, what was the metallicity of the gas from which Pop. III stars formed? (Only the lightest three elements, H, He and Li, were produced in the big bang. The fraction of helium produced by the big bang was 0.075 of the total number of nuclei, and each helium nucleus has a mass four times that of hydrogen.)

SOLUTION

Since there are only three elements (H, He and Li) produced in the big bang, we can begin with the definition for metallicity (Equation 1.1)

$$Z = \frac{\text{the mass of elements heavier than helium in the gas}}{\text{the mass of all elements in the gas}}$$

and rewrite it as

$$Z = \frac{\text{the mass of lithium in the gas}}{\text{the mass of all elements in the gas}}$$

We introduce the symbols N_X for the number of nuclei of element X, and M_X for the mass of each nucleus of element X, giving

$$Z = \frac{N_{Li}M_{Li}}{N_H M_H + N_{He}M_{He} + N_{Li}M_{Li}}$$

We don't know the total number of nuclei $N_{tot} = N_H + N_{He} + N_{Li}$ produced in the big bang, but that doesn't matter because we know what *fraction* of the total is accounted for by each type of nucleus. That is, we know the values N_H/N_{tot}, N_{He}/N_{tot}, and N_{Li}/N_{tot}. (These were given in the question, and come from astronomical observations.) To proceed, we divide the top and bottom lines by N_{tot}, and also by M_H for reasons that will become clear.

$$\begin{aligned} Z &= \frac{(N_{Li}M_{Li})/(N_{tot}M_H)}{(N_H M_H + N_{He}M_{He} + N_{Li}M_{Li})/(N_{tot}M_H)} \\ &= \frac{(N_{Li}/N_{tot})(M_{Li}/M_H)}{(N_H/N_{tot})(M_H/M_H) + (N_{He}/N_{tot})(M_{He}/M_H) + (N_{Li}/N_{tot})(M_{Li}/M_H)} \end{aligned}$$

This can now be evaluated, because all the ratios that appear in the equation are given in the question. Note that, since only H, He, and Li are produced in the big bang, the number of hydrogen nuclei is given by $N_H = N_{tot} - N_{He} - N_{Li}$, so $N_H/N_{tot} = 1 - (N_{He}/N_{tot}) - (N_{Li}/N_{tot})$, so

$$\begin{aligned} Z &= \frac{1.6\times10^{-10}\times7}{(1-0.075-1.6\times10^{-10})\times1+0.075\times4+1.6\times10^{-10}\times7} \\ &= \frac{1.12\times10^{-9}}{0.925+0.30+1.12\times10^{-9}} = 9.1\times10^{-10} \\ &\approx 10^{-9} \end{aligned}$$

That is, the metallicity of a Pop. III star is expected to have been $Z \sim 10^{-9}$.

QUESTION 1.2

(a) Describe Baade's observations of M31 and their implications for the Milky Way. (*Note*: Baade did not know the details of stellar evolution – those did not become clear until many years later – so your answer should *not* discuss stellar evolution.)

(b) From what you know about the evolution of stars, how would you interpret the differences between red and blue stars in Baade's observations?

QUESTION 1.3

Stellar populations differ in age, metallicity and location, but sometimes stellar motion is used as an alternative criterion to location. Why is this reasonable?

1.2.5 The chemical evolution of the Milky Way

We have seen that the stellar populations of the Galaxy have a range of ages, metallicities and locations. To understand why the different populations have different ranges of metallicity we have to consider where the metals come from.

The only metal produced in the big bang was lithium, all others result directly or indirectly from nuclear reactions occurring in stars. The metallicity of a main sequence star corresponds to the metallicity of the ISM at the time the star formed. The fact that each stellar population contains a range of metallicities, and that the younger stars in a population tend to have higher metallicities, therefore suggests the operation of some process that progressively enriches the ISM by increasing its metallicity. There are a number of routes by which metals that form inside a star can escape and enter the surrounding ISM, thereby bringing about the required chemical enrichment. These escape routes only become available during the late stages of a star's life (although these come relatively quickly for high-mass stars), and may include the emission of high-speed stellar winds, the ejection of shells of gas to form planetary nebulae, and, in some cases, disruption of the star in the explosive process known as a supernova. By whichever routes it happens, the transfer of chemically enriched material from a star to the ISM enriches the ISM and ensures that the next generation of main sequence stars to form in that region will have higher metallicity than its predecessor. Thus the chemical evolution of the Galaxy is a cyclic process involving star formation, element production within stars (nucleosynthesis), and the return of chemically enriched material to the ISM where it can form more stars. This process is sometimes called **cosmic recycling**, and is central to our understanding of the differences between Pop. I and Pop. II stars in terms of the way the Galaxy has evolved since its formation.

Although cosmic recycling will have taken place in each of the Galactic components (disc, bulge and stellar halo) that contain stars, it probably did not proceed at the same rate in each of them, nor does it necessarily continue at a significant rate in each of them today. The presence of various amounts of metals in stars of different ages allows astronomers to deduce the history of cosmic recycling in the various stellar populations, which in turn allows them to trace the star formation history of the Galaxy. We will explore this topic in more detail later, particularly in Section 1.6.2.

Before concluding this section we provide a few questions to encourage you to think about the link between the content of the Galaxy and its evolution.

QUESTION 1.4

Why would you expect surviving Pop. II main sequence stars to have lower metallicity than Pop. I main sequence stars?

QUESTION 1.5

Some people think that old stars have been undergoing nucleosynthesis for a long time, so when we observe Pop. II main sequence stars they should exhibit higher metallicity than the younger Pop. I stars because of the accumulated products of nucleosynthesis. Explain why this view is wrong.

1.3 The mass of the Milky Way

When the structural components of the Galaxy were introduced in Section 1.2.1, their masses were simply stated. At the time, did you ask yourself how astronomers might know these values, or how uncertain they might be? You may be surprised to learn that reasonable estimates can be obtained from some quite simple calculations based on the influence of gravity on the motions of objects, although determining the mass of the entire Galaxy accurately is a major challenge.

In answering Question 1.1, you learned that the Sun takes roughly 240 million years to complete one orbit of the Galactic centre. To perform this calculation you needed to know the speed at which the Sun travels in its orbit and its distance from the Galactic centre. The speed of the Sun was established in the 1920s, by Bertil Lindblad (1895–1965) and Jan Hendrik Oort (1900–1992), from the motion of Pop. I stars relative to Pop. II stars. As you saw in Section 1.2.4, Pop. II stars have little net rotation about the Galactic centre, and so provide a reference population relative to which the Sun's motion can be determined. Pop. I stars in our part of the Milky Way are streaming past the Pop. II stars at speeds that are typically 220 km s^{-1}. We will soon use this information to calculate the mass of the inner part of the Galaxy, but before doing so we introduce a graphical tool that plays an important part in the characterization and analysis of rotating systems.

A plot of speed against distance from the centre for the various parts of a rotating system is called a **rotation curve**. The rotation curve for a rigid wheel 3 m in diameter, making one revolution per second, is displayed in Figure 1.13a. As the figure shows, the speed (usually expressed in metres per second; m s^{-1}) of each part of the wheel increases in proportion to the distance from the centre. Note, however, that in the case of a rigid wheel the angular speed of each part about the centre – the angle that each part sweeps out per second, as viewed from the centre of the wheel – is the same. Finding that all parts of a body have the same angular speed is a characteristic of **rigid body rotation**.

Figure 1.13b shows the rotation curve of planets orbiting the Sun. Each planet takes a different length of time to complete one orbit, which means each travels at a different angular speed. This is characteristic of **differential rotation**, and is clearly different from rigid body rotation. As you can see, the rotation curve for such a system is *not* a straight line through the origin, so speed is *not* proportional to distance from the centre.

Figure 1.13c shows the rotation curve of the Milky Way. This has a different shape from Figure 1.13b, but it still shows that the speed is not proportional to distance from the centre, and therefore indicates differential rotation with a range of angular speeds. The difference between Figures 1.13b and 1.13c is due to the fact that there is no massive central body dominating the Milky Way, whereas the mass of the Sun dominates the Solar System.

1.3.1 Calculating the mass of a gravitating system

The equations describing the rotation curve of a system governed by the gravitational attraction of a single central body relate the orbital speed and orbital radius to the mass of the central body. This is very important in astronomy, as it allows us to calculate the mass of a gravitating body from the motions of particles orbiting it. For an object in a circular orbit of radius r about a single,

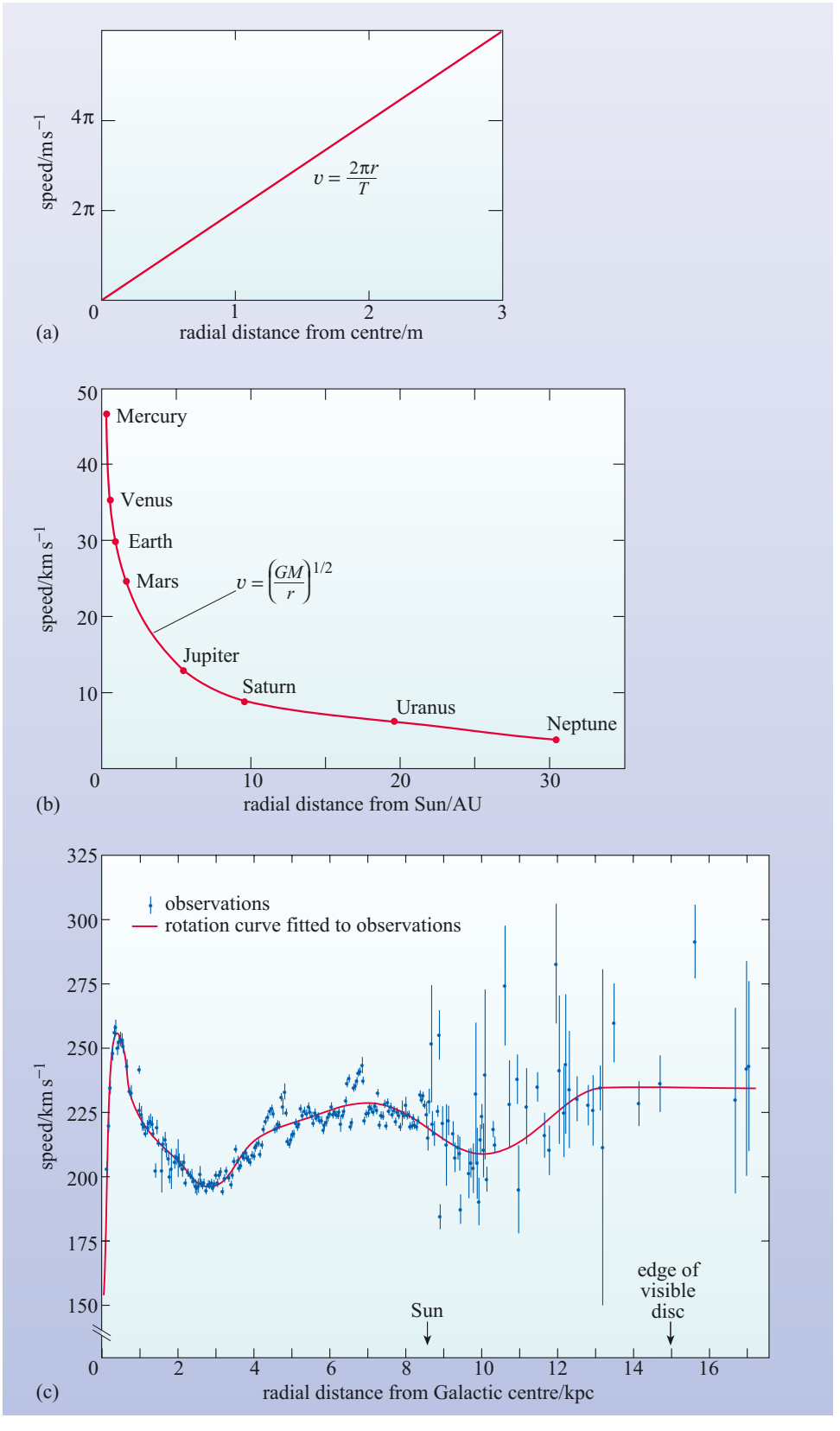


Figure 1.13 The rotation curves, showing speed v against radial distance r, of (a) a rigid wheel, 3 m in diameter, with a rotation period T of 1 s; (b) the planets of the Solar System; and (c) the Milky Way, based on Doppler-shift studies of gas clouds in the disc. Note that the rotation curve for the Galaxy is based on data that are subject to substantial observational uncertainties, and is therefore 'noisier' than the other two curves. This is because the observations are harder to make (especially for objects further from the Galactic centre than the Sun), and the analyses are plagued by additional complications, such as the non-uniform distribution of matter in the Galactic disc. ((c) Combes, 1991)

much more massive object of mass M, the equation describing the rotation curve (see Box 1.2) is

$$v = \left(\frac{GM}{r}\right)^{1/2}$$

The reciprocal of some quantity x is $1/x$.

The symbol $\propto$ means 'is proportional to'.

Note that $\sqrt{x} = x^{1/2}$

This equation indicates that the speed v is proportional to $1/r^{1/2}$, which means that the orbital speed falls as the radius increases, at a rate given by the reciprocal of the square root of the radius. We can write this in mathematical shorthand as $v \propto 1/r^{1/2}$. The curve plotted in Figure 1.13b has this functional form.

BOX 1.2 ROTATION CURVES FOR GRAVITATING SYSTEMS

According to Newton's second law of motion, the magnitude of the acceleration a_i of some body (labelled 'i') of mass m_i, due to a force of strength F, is

$$a_i = F/m_i \qquad (1.2)$$

According to Newton's law of gravitation, the strength F_g of the gravitational force on each of two point-like bodies of masses m and M, when their centres are separated by a distance r, is

$$F_g = GMm/r^2 \qquad (1.3)$$

where G is the universal gravitational constant, $6.673 \times 10^{-11}\,\mathrm{N\,m^2\,kg^{-2}}$.

Substituting this expression for F_g into Equation 1.2 shows that, due to the gravitational attraction of the body of mass M, the body of mass m will have an acceleration of magnitude

$$a_{g,m} = \frac{GMm/r^2}{m} = GM/r^2$$

Similarly, due to the attraction of the body of mass m, the body of mass M will have an acceleration of magnitude

$$a_{g,M} = \frac{GMm/r^2}{M} = Gm/r^2$$

In cases where M greatly exceeds m, written $M \gg m$, we can also write $GM/r^2 \gg Gm/r^2$, implying that $a_{g,m} \gg a_{g,M}$. That is, the acceleration of the more massive body has a magnitude, $a_{g,M}$, that is much smaller than the magnitude of the acceleration of the less massive body, $a_{g,m}$. Therefore the more massive body barely moves, and we can regard it as being the stationary centre of motion for the less massive body. One particularly simple form that this motion might take is for the less massive body to move around the more massive body at constant speed in a circular orbit. This kind of motion is known as uniform circular motion.

Any body moving in uniform circular motion at speed v about some centre at a distance r must be accelerating at all times. This acceleration is called its centripetal acceleration; it is always directed towards the centre of the motion and its magnitude is

$$a_{cen} = v^2/r$$

When this uniform circular motion is the result of the gravitational attraction between two bodies, the centripetal acceleration is provided by the gravitational acceleration, so for the less massive body in motion about the stationary, more massive one, we can write:

$$a_{cen} = a_{g,m}$$

so $\quad v^2/r = GM/r^2$

and hence

$$v^2 = GM/r$$

This equation can be rearranged in two slightly different ways to give useful equations relating the orbital speed and the central mass of a two-body system:

$$v = (GM/r)^{1/2} \qquad (1.4)$$

and

$$M = v^2 r/G \qquad (1.5)$$

Note that the orbital speed of the less massive body does not depend on its mass. Also notice that we can deduce the mass of the central body from the speed and radial separation of an orbiting body, without knowing the mass of that orbiting body.

QUESTION 1.6

(a) Calculate the circumference of the Earth's orbit around the Sun. (The Earth is 150 million kilometres from the Sun, to three significant figures.) Give your answer in SI units.

(b) Calculate the speed at which the Earth orbits the Sun. Give your answer in SI units.

(c) Use the formula for the rotation curve to calculate the mass of the Sun from the orbital speed of the Earth. Give your answer in SI units. (For the universal gravitational constant, use the value $G = 6.673 \times 10^{-11}\,\mathrm{N\,m^2\,kg^{-2}}$.)

(d) How many significant figures are there in your final answer, and why?

How does your answer to Question 1.6 compare with the modern value, $M_\odot = 1.9891 \times 10^{30}\,\mathrm{kg}$? Did you ever realize just how easy it would be to calculate the mass of the Sun? All you need to know is the speed and orbital radius of one small body in a circular orbit. You don't even need to know the smaller body's mass.

You could look up the orbital radii and periods of a few other planets to convince yourself that you get the same answer, but for now we have bigger fish to fry; let's calculate the mass of the inner part of the Galaxy!

Unlike the Solar System, the Milky Way does not have a single dominant mass at its centre. Rather, its constituents move under the gravitational influence of all the other constituents. This makes a detailed analysis very complicated, but it is still possible to make a simple estimate of the mass of the inner part of the Galaxy. The basis of this analysis is provided by the following result taken from Newtonian gravitation theory.

> When the mass of a system is distributed in a spherically symmetrical manner about some central point, then the net gravitational force on a point-like object at some radius is due only to the mass *within* that radius. Furthermore, the net gravitational force is the same as if the mass inside that radius was all located at the centre.

In the case of the Milky Way, the distribution is not really spherically symmetrical, but mass outside a given radius has only a moderate effect, and a reasonable estimate of the mass can still be obtained using the method above. (Detailed calculations can be performed for more realistic mass distributions to confirm this.) If we adopt this procedure for the Galaxy, then the observed orbital speed at some radius r can be inserted into the rotation-curve equation to estimate the mass of the Galaxy enclosed within that radius from the Galactic centre. As this gives the mass within the radius r, rather than the total mass of the Galaxy, we denote that mass by $M(r)$.

QUESTION 1.7

(a) Following the technique used in Question 1.6c, calculate the mass of the Milky Way out to the distance of the Sun from the Galactic centre. Give your answer both in SI units and solar masses, assuming $M_\odot = 1.99 \times 10^{30}\,\mathrm{kg}$. (*Hint*: You know already that the Sun is 8.5 kpc from the Galactic centre, and that it moves at $220\,\mathrm{km\,s^{-1}}$ in a nearly circular orbit.)

(b) How many significant figures can you quote the result to, if you follow the usual mathematical rules? Is there any physical reason why you might deviate from this rule?

1.3.2 Using rotation curves

The answer to Question 1.7 shows that interior to the Sun's orbit at 8.5 kpc, the mass of the Galaxy is $10^{11} M_{\odot}$. If we want to find the *total* mass of the Milky Way, we have to study its outskirts, where the orbiting material encloses virtually all the mass of the Galaxy. This is difficult, not least because it is difficult to determine exactly where the Galaxy ends. Even if we think there is not much more *visible* matter beyond a certain radius, we cannot be sure that we have found the 'edge' of any dark matter that is associated with the Galaxy.

Plotting a rotation curve can throw light on the question: 'Where does the Galaxy end?' We have just seen in Question 1.7 how it is possible to compute the enclosed mass at some radius in the Galaxy by knowing the orbital speed there. From an observed rotation curve, it is possible to compute the enclosed mass $M(r)$ at a whole range of radii, and doing so shows how $M(r)$ increases with radius, which therefore gives the *distribution* of mass. By seeing how the mass distribution is changing in the outermost *measurable* parts of a Galaxy, it is possible to have some idea of whether the mass distribution is tailing off near the last measurement.

The usual procedure for deducing $M(r)$ for a galaxy is to take an educated guess at the distribution of matter and then work out the rotation curve that such a distribution would produce. The initial guess is then adjusted until the modelled rotation curve agrees with the observed one. To see how this process works, in the following example and question we compute the rotation curves for some simple, assumed mass distributions, and then use these results to interpret what has been measured for the Milky Way, which is shown in Figure 1.13c.

EXAMPLE 1.2

Use the rotation-curve equation to help you to sketch a rotation curve for the following distribution of matter: $M(r) = M$, a constant, indicating a central mass only.

SOLUTION

A rotation curve is a plot of speed versus radius, so the more useful form of the rotation-curve equation (see Box 1.2) is $v(r) = (GM(r)/r)^{1/2}$ (Equation 1.4). To sketch the rotation curve, we need to know how v varies with r.

Since in the example $M(r)$ is a constant, M, the equation for the speed becomes

$$v(r) = (GM/r)^{1/2} = \text{const} \times 1/r^{1/2}$$

Note that we have put the constant parts of this expression into one term called 'const', and have kept the variable parts separate.

This equation shows that the rotation curve falls with increasing radius as $1/\sqrt{r}$, so your sketch of the rotation curve in this case should decrease in speed as radius increases, and it should flatten out towards large radius. One example of a system dominated by a central mass is the Solar System, so the rotation curve should match Figure 1.13b.

Historically, it was Johannes Kepler (1571–1630) who first recognized that a relationship of this form describes the motion of planets in the Solar System. This is the origin of the term **Keplerian orbit** that astronomers now use to refer to the motion of a body under the gravitational influence of a much more massive body.

QUESTION 1.8

Following the example presented above, use the rotation-curve equation to help you to sketch a rotation curve for each of the following distributions of matter.

(a) $M(r) = kr$ (for some constant of proportionality k);

(b) a uniform-density sphere, i.e. where $M(r)$ = density × (volume of sphere of radius r) = $\rho \times \frac{4}{3}\pi r^3$

Figure 1.13c shows the measured rotation curve of the Milky Way. Some of the features of Figure 1.13c, such as the peak near the Galactic centre and the sharp dip that follows it, are more likely to be due to the inadequacy of the symmetry assumptions that underpin the analysis rather than real features of the rotation. However, the flatness of the rotation curve at large distances is thought to be real. It is flatness of this kind, extending well beyond the edge of the visible disc, which provides evidence for the presence of a substantial amount of non-luminous matter on the outskirts of the Milky Way – dark matter. If you compare all but the central parsec of the Milky Way's rotation curve (Figure 1.13c) with your answers to Question 1.8, you will see that it is similar to the flat curve for $M(r) = kr$. Note that a mass distribution of the form $M(r) \propto r$ thins out with increasing radius, so a flat rotation curve does not imply constant density.

If most of the mass of the Galaxy were well within the largest measured radius, the rotation curve would be expected to decline quite rapidly with increasing radial distance, as in the case $M(r)$ = constant in the solution to Example 1.2. The answer to the question: 'What is the mass of the Galaxy?' really depends on the answer to another question: 'Where does the rotation curve turn down?'

Several independent investigations have failed to show any sign of a decline in the rotation curve of the Milky Way out to a radius of 20 kpc, indicating a substantial amount of matter at least out to that radius. This matter has not been directly observed at any wavelength and is therefore some form of dark matter.

There are still many uncertainties about the distribution of mass in the Milky Way. Different assumptions about the radius of the Galaxy and the distribution of dark matter can easily provide estimates of the total Galactic mass that range from a conservative four times the mass of the stars, that is $4 \times 10^{11} M_\odot$, to a very substantial 60 times: $6 \times 10^{12} M_\odot$. The mass of the Galaxy can be assessed using the velocities of other objects besides disc gas, such as distant halo stars, globular clusters and nearby galaxies, but they too are based on specific assumptions and do not yet settle the question.

We noted at the beginning of this chapter that measuring the *size* of the Galaxy was not as simple a task as it might first sound. Astronomers Michel Fich and Scott Tremaine described the problem of measuring the *mass* of the Galaxy by attacking the question, in the following way:

> 'What is the mass of the Galaxy? The most important recent progress in addressing this question has been the recognition that it is not well-posed. … there is probably no natural definition of the mass of a giant galaxy like our own …'

■ Could you calculate the mass of the wheel whose rotation curve is shown in Figure 1.13a, using the technique used in Question 1.7? Would it make any difference if most of the mass of the wheel were concentrated near the centre of rotation? Explain your answer.

❑ You could not determine the mass of the wheel; the rotation curve contains no information on its mass because rotation of the wheel is governed by non-gravitational forces. (The forces are electrostatic, and act between the many neighbouring particles in the rigid structures of the wheel.) You could calculate the centripetal acceleration of various parts of the wheel, but those accelerations would not be related to its mass distribution. This answer would not be any different if the mass were concentrated near the axle.

1.4 The disc of the Milky Way

The major components of the Galaxy have now been introduced, their chief constituents have been described, and, at least as far as its inner parts are concerned, the mass of the Galaxy has been estimated. In the remaining sections of this chapter we examine the disc, the stellar halo and the bulge in more detail, and we begin to uncover the evolutionary history of the Milky Way. We start in this section with a detailed account of the Galactic disc.

We have seen already that dark matter dominates the mass of the Galaxy as a whole, but if we consider the disc alone, we find that there the dark matter has less impact. The motion of stars perpendicular to the disc indicates that no more than 30–50% of the *disc's* mass is due to dark matter. Stars and gas are so abundant in the disc that visible matter dominates there.

The disc is important to our understanding of the Galaxy for two reasons: it is the component to which most of the visible matter – stars, gas and dust – belongs, and it is the main site of current star formation in the Milky Way. Most of that star forming

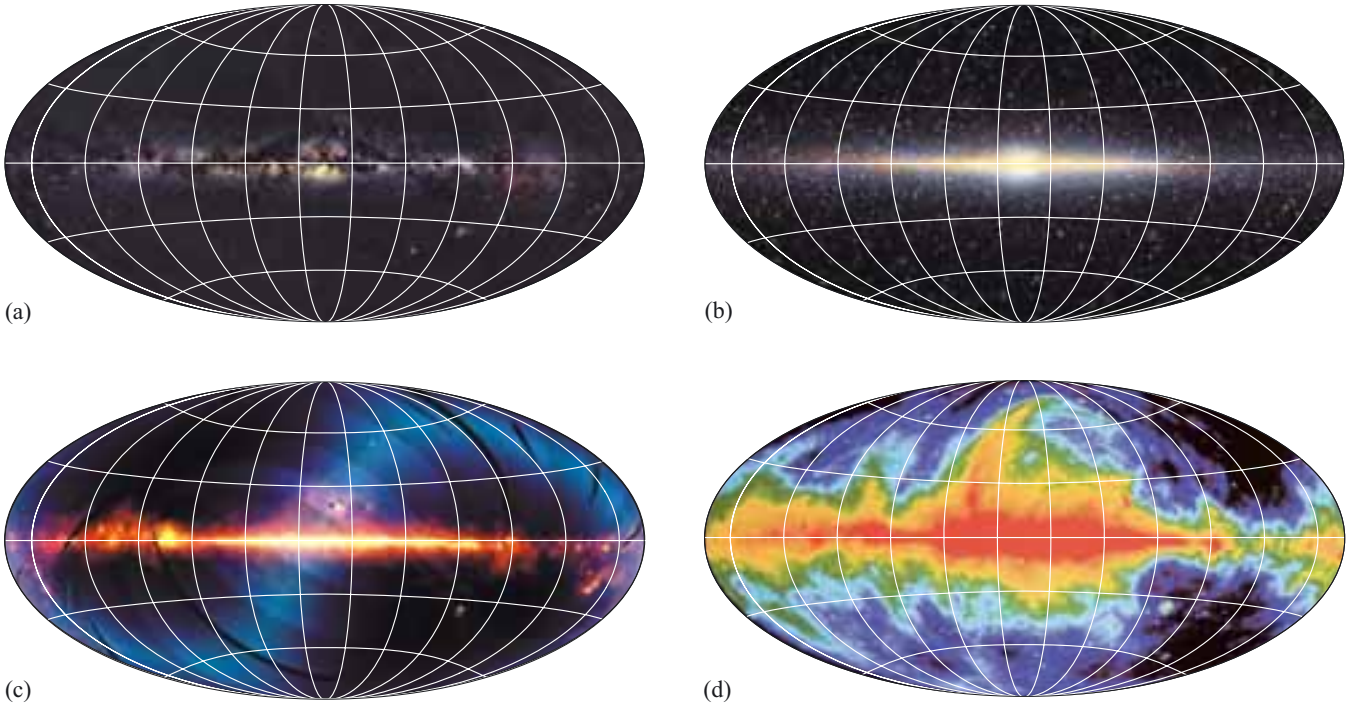


Figure 1.14 All-sky views of the Milky Way at various wavelengths. The symmetry of the Galaxy and the proximity of the Sun to the mid-plane of the disc are emphasized in these views, which are presented in Galactic coordinates with the Galactic equator running horizontally across the middle. They are arranged with the Galactic centre ($l \approx 0°$, $b \approx 0°$) near the middle. (a) Optical light which is predominantly the thermal radiation from stars; (b) near-infrared emission dominated by thermal radiation from cool stars; (c) far-infrared emission dominated by thermal radiation from dust, especially that associated with star-forming regions; (d) 73.5 cm radio continuum emission. Note that the blue 'S'-shaped band in the far-infrared image (c) is due to hot dust within our Solar System. ((a) A. Mellinger (University of Potsdam, Germany); (b) N. Wright; (c) NASA/Goddard Space Flight Center; (d) G. Haslam *et al.*, (Max-Planck-Institut für Radioastronomie, Bonn))

activity occurs in the spiral arms, which would be a prominent feature of the disc if we could view it externally. Many of the characteristic features of the disc that we can observe, such as the presence of young, high-metallicity stars, dense molecular clouds and HII regions, are directly connected with star formation. A major aim of this section is to survey these observable features and to relate them to star formation and cosmic recycling.

The visual appearance of the Milky Way (Figure 1.14a) is dominated by luminous stars, mostly belonging to the disc, although in various directions the view is obscured by dark dust clouds. Viewed at the longer wavelengths of near-infrared radiation (Figure 1.14b), the dust clouds become transparent, allowing us a relatively clear line of sight towards the Galactic centre. At the still longer wavelengths of far-infrared radiation, the dust itself becomes a luminous source of radiation (Figure 1.14c), and at even longer radio wavelengths the Galactic gas that surrounds us becomes visible (Figure 1.14d), creating a view that is very different from that of the visible stars.

- On the maps in Figure 1.14, where is the Galactic anti-centre, i.e. the direction directly away from the Galactic centre?

- It is the point that lies on the Galactic equator, 180° from the Galactic centre. This point appears both at the far left and far right ends of each map, where they 'wrap around'.

1.4.1 The stellar content of the disc

Stars are the most luminous constituent of the Milky Way, and most stars reside in the Galactic disc. Usually stars do not form alone, but instead begin their lives in clusters or associations. This is a consequence of stars forming from dense molecular clouds (see Section 1.2.3) that contain enough gas to create large numbers of stars. **Open clusters** occupy regions of space, typically 2–3 pc across, where the density of stars is enhanced locally by a group of a few tens to a few hundred stars that formed at the same time. The cluster members are bound together by their mutual gravitational attraction. There are thousands of open clusters in the Galactic disc. Some are sufficiently prominent to be visible to the naked eye, most notably the Pleiades (Figure 1.15).

Because of their concentration close to the plane of the Galaxy, open clusters used to be called *galactic clusters*. However, this term is potentially confusing ('galaxy clusters', which is a term sometimes used to describe clusters *of* galaxies, are quite different) and its use is discouraged.

Figure 1.15 The Pleiades is a very young open cluster with an age of only ~80 Myr. Massive hot, blue stars dominate its appearance. (D. Malin/Royal Observatory Edinburgh/Anglo-Australian Telescope Board)

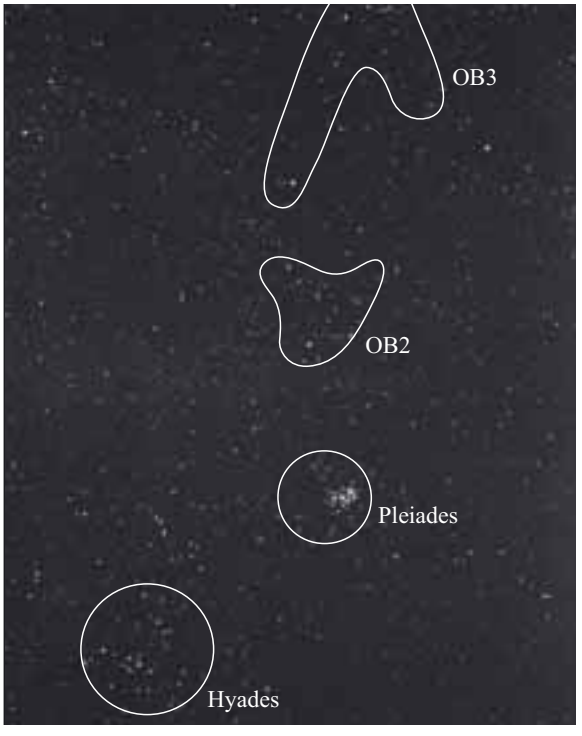


Figure 1.16 A portion of the Milky Way showing the Hyades and Pleiades open clusters, and OB associations in Taurus and Perseus.

Open clusters are believed to have relatively short lives. With very few exceptions no individual open cluster is expected to survive for more than 10^9 years. The smallest open clusters live for even shorter periods, just a few million years, so they are mainly found in the spiral arms, since this is where star formation mainly takes place. Why do the clusters disrupt and dissipate in a time that is much shorter than the lives of many of the stars they contain? There are three processes that aid this. First, a cluster moving in the disc might be torn apart by a gravitational encounter with some other object such as a giant molecular cloud or another cluster. Secondly, gravitational interactions between a cluster's members can give one star enough energy to escape; given long enough, this process, called **evaporation**, would disperse all clusters. A third process that disrupts clusters is differential rotation, which was described in Section 1.3. The stars in an open cluster, while moving around one another, are also orbiting the Galactic centre and hence are subject to the effects of differential rotation that will stretch, distort and eventually destroy the cluster. The concentration of open clusters in the spiral arms must mean that the clusters have relatively short lives, otherwise motion relative to the arms would relocate them to various positions in the disc.

Stellar aggregates of another kind found in the disc, particularly within the spiral arms, are the so-called **OB associations**. OB associations have diameters of 100 pc or so, and densities not much greater than their general surroundings, but contain an unusually high proportion of O- and B-class stars (Figure 1.16). Many OB associations have a very young open cluster at their centre. The presence of numerous O and B stars, which have very short main sequence lifetimes, is a sure sign that OB associations are young, no more than a few million years old. About 70 OB associations are known.

The stars of the disc are primarily Pop. I objects. They can be divided into a range of subpopulations, three of which are described below. Other schemes can be used too; the important thing is not to get hung up on these definitions, but rather to appreciate that a range of objects exists and that *sometimes* it is useful to make distinctions between them.

- *Pop. I, spiral-arm stars* are the youngest stars in the Galaxy, with ages less than about 0.1×10^9 yr, and are found in the spiral arms. They are associated with the spiral arms because that is where they have formed, and their short lives have not enabled them to move far from their birthplace. Examples include stars in young open clusters such as the Pleiades (Figure 1.15) and the Hyades (Figure 1.16), short-lived massive stars such as O and B stars, supergiant stars, and the pulsating giant stars called **classical Cepheids**. The young stellar objects known as **T Tauri stars** also belong to this subpopulation. The metallicities of Pop. I spiral-arm objects are typically solar or greater, i.e. $Z = 0.02$ to 0.04. Associated objects include glowing HII regions (see Section 1.2.3), which are produced where high-energy UV photons from O- and B-type stars ionize hydrogen gas in the interstellar medium.
- *Pop. I, thin-disc stars* are older than the spiral arm stars, with ages from (1 to almost 10) $\times 10^9$ yr, and their metallicities include some much lower values, $Z = 0.005$ to 0.04. These stars move in circular orbits, and have lived long

enough to escape from the spiral arms and spread themselves across the disc. The word 'thin' is included in the name because they are seldom found more than 500 pc from the mid-plane of the disc, so they occupy a **thin disc** with a relatively small cross-section. The significance of this will become clearer when we discuss the next subpopulation, the thick-disc stars.

- *Intermediate Population or thick-disc stars* have lower metallicities than thin-disc stars, typically $Z \approx 0.002$ to 0.01, and ages closer to the thin-disc maximum, 10×10^9 yr. Another distinguishing feature is that while their orbits are still basically circular, the thick-disc stars travel to greater distances from the Galactic plane. This last characteristic means they are not confined as close to the Galactic plane as normal thin-disc stars, so they occupy a **thick disc** with a greater cross-sectional area than the thin disc. Their typical locations and metallicities give these objects properties intermediate between those of the Pop. I and Pop. II stars discussed earlier, so they are also known as **Intermediate Population** stars.

In Section 1.2.5 we introduced the concept of cosmic recycling, and discussed the differences between Pop. I and Pop. II stars in terms of the chemical evolution of the Galaxy. The existence of the three subpopulations described above now emphasizes that there is also an evolutionary sequence within the disc, with thick-disc stars having formed 10×10^9 yr ago from low metallicity gas, and spiral-arm objects having formed much more recently, less than 0.1×10^9 yr ago, from gas with a much higher metallicity. The existence of stars with a range of ages, metallicities and motions provides clues to the evolutionary history of the disc, and indicates that it is continuing to evolve. We return to this topic in Section 1.6.

■ Why would you expect HII regions to be associated with the spiral arms?

❑ HII regions need the presence of O and B stars to provide ionizing UV photons, and O and B stars are themselves associated with the spiral arms because they are short-lived stars that do not survive long enough to move very far away from their sites of formation.

1.4.2 The gaseous content of the disc

The gaseous interstellar medium (ISM) is intimately associated with stellar evolution. Stars form from cool dense clouds in the ISM, and at the ends of their lives they return matter to the ISM in a variety of ways that gradually increase the metallicity of the Galaxy. Gas and dust are therefore important constituents of the disc, and we examine their composition and distribution in this section. Included in the discussion are descriptions of how the gas can be observed.

Interstellar gas

As you saw earlier, the total mass of gas in the disc amounts to about 10% of the stellar mass. The gas forms a disc about 300 pc thick that is surrounded by hotter gas, which stretches out into the halo. Much of the gas is in the form of *clouds*, of which there are many thousands, with a range of temperatures and densities. The individual clouds probably take the form of sheets or filaments of gas rather than some idealized spherical shape. Regions of the ISM not occupied by clouds constitute the extensive but low-density *intercloud medium*. There the hydrogen exists mainly in the form of neutral atoms (HI), which are easily observed due to their emission of radio waves at a wavelength of 21 cm (see Box 1.3).

BOX 1.3 THE 21 CENTIMETRE EMISSION LINE OF ATOMIC HYDROGEN

A major indicator of the distribution and line-of-sight velocity of neutral (atomic) hydrogen (HI) gas, not just in the Milky Way but other galaxies as well, is its emission line in the radio region of the spectrum at a wavelength of 21 cm. The origin of this **21 centimetre radiation** involves the relative *spins* of the electron and proton that constitute the hydrogen atom. These spins are illustrated in a classical (i.e. non-quantum) sense in Figure 1.17, where the proton and the electron are pictured as small spheres spinning at a fixed rate around axes through their centres. The quantum physics of the hydrogen atom ensures that the electron spin is always either parallel to that of the proton, as in Figure 1.17a, or anti-parallel (i.e. opposed to it), as in Figure 1.17b. There is a small energy difference between the states shown in Figure 1.17, and it is the transition from the higher energy state (a) to the lower energy state (b) that gives rise to the 21 cm emission line.

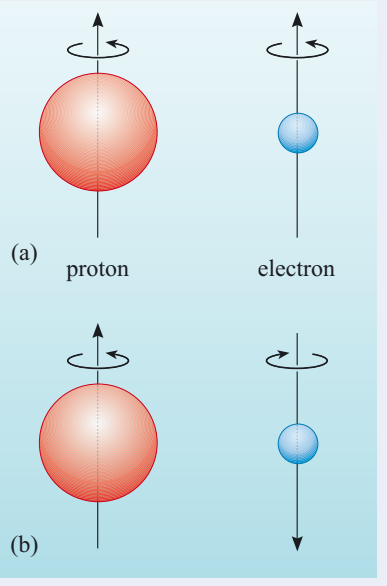


Figure 1.17
A classical view of the hydrogen atom, in which the proton and electron spins are (a) parallel or (b) anti-parallel.

■ What is the energy difference between these two states, in SI units (joules) and electronvolts?

❑ The energy difference between the states is just the energy of the emitted photon. This is given by

$$\varepsilon = hf = hc/\lambda \qquad (1.6)$$

Thus the energy difference between the two states associated with the 21 cm line is

$$\begin{aligned}\varepsilon &= hc/\lambda \\ &= 6.63 \times 10^{-34}\,\text{J s} \times 3.00 \times 10^{8}\,\text{m s}^{-1}/0.21\,\text{m} \\ &= 9.5 \times 10^{-25}\,\text{J}\end{aligned}$$

Since $1\,\text{eV} = 1.60 \times 10^{-19}\,\text{J}$, the energy difference can also be expressed as

$$\begin{aligned}\varepsilon &= 9.5 \times 10^{-25}\,\text{J}/1.60 \times 10^{-19}\,\text{J eV}^{-1} \\ &= 5.9 \times 10^{-6}\,\text{eV}\end{aligned}$$

For us to observe this emission line, the hydrogen atoms must be in an environment where they can readily gain the energy required to raise them into the upper energy level at a reasonable rate compared with the rate at which they are reverting to the lower energy level by emitting radiation. One energy source is provided by collisions between hydrogen atoms as a result of their random thermal motion – this is an example of **collisional excitation**. For a reasonable proportion of such collisions to be sufficiently energetic, the average translational kinetic energy of an atom, e_k, must exceed the energy difference between levels, ε, such that $e_k \geq \varepsilon$. For thermal motion, the average translational kinetic energy is related to the temperature T of the gas particles via the equation $e_k = 3kT/2$, where k is the Boltzmann constant. Thus, by requiring $e_k \geq \varepsilon$, we get $3kT/2 \geq \varepsilon$ and hence $T \geq 2\varepsilon/3k$.

Putting in the value of ε, we get the requirement $T \geq 2 \times 9.5 \times 10^{-25}\,\text{J}/(3 \times 1.38 \times 10^{-23}\,\text{J K}^{-1}) = 0.046\,\text{K}$. This condition is met everywhere in the ISM, so the 21 cm line is readily emitted wherever atomic hydrogen exists.

The radial velocity of a cloud of atomic hydrogen can be measured from the Doppler shift of the 21 cm emission line. In general, the radial velocity (i.e. the component of velocity along the line of sight) of an object that emits (or absorbs) radiation at a wavelength λ_{em} is given by

$$v_r = c(\lambda_{obs} - \lambda_{em})/\lambda_{em} \qquad (1.7)$$

where λ_{obs} is the wavelength at which the radiation is observed and c is the speed of light. This relationship is valid provided that the radial velocity is much smaller than the speed of light, v_r must be less than about $0.1c$. Note also that the convention used in Equation 1.7, and throughout this book, is that an object moving *away* from the observer has a *positive* radial velocity.

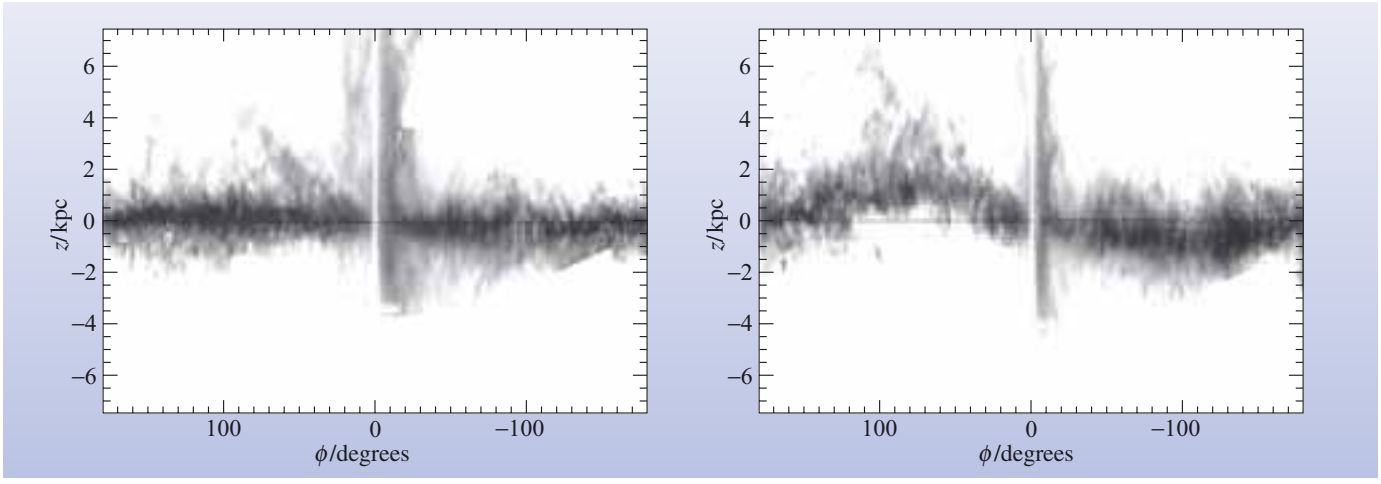


Figure 1.18 The Galactic HI density on cylindrical surfaces at (a) 12 kpc and (b) 16 kpc from the Galactic centre. Each map plots the distribution of gas at a vertical displacement, z, from the Galactic plane and an azimuthal angle, ϕ, measured in the Galactic plane. Each map is constructed as if viewed from the Galactic centre. The warping of the disc, which sets in around 16 kpc, is seen as a wave-like displacement of the gas from the equatorial plane, strongest around $\phi = \pm 90°$. (The vertical column of gas in the centre of each image is an artefact of the way the maps have been made.) (Binney and Merrifield, 1998; from data published in Burton, 1985; Hartmann and Burton, 1997; and Kerr *et al*., 1986, courtesy of T. Voskes and B. Burton)

Although we have said that the gas forms a disc with a thickness of 300 pc, this disc is not completely flat. Its mid-plane is flat out to a radius of 12 kpc from the Galactic centre, but at greater distances it is warped (tilted). This can be seen in the two parts of Figure 1.18, which show the gas distribution at 12 kpc and 16 kpc from the Galactic centre, as it would be seen from the Galactic centre rather than from the location of the Solar System. The images are built up from measurements of the hydrogen 21 cm emission line. The figure shows that at a distance of 12 kpc, the gas distribution is centred about the Galactic plane, indicating that the disc is flat at this distance from the Galactic centre. However, at 16 kpc the gas rises above and falls below the Galactic plane (see Figure 1.19). This demonstrates the presence of a warp, which can be interpreted as a tilt of the gaseous disc at distances around 16 kpc. The tilt is also present beyond 16 kpc. The gas reaches an altitude (z) of 1 kpc to 2 kpc above the plane in one azimuthal direction ($\phi = +90°$), and a similar distance below the plane in the opposite direction, ($\phi = -90°$). Roughly one-quarter of galaxies have warped discs, and although various possible causes of warping have been explored, no compelling explanation has been developed.

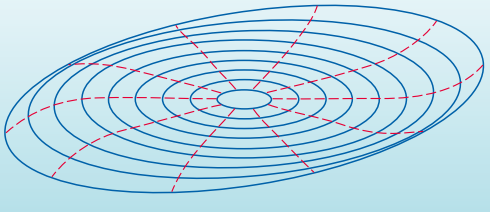


Figure 1.19 A schematic diagram of the warped disc of gas. Nine circular contours 2 kpc apart are shown. The inner six contours of the disc are in the same plane, but the outer three are progressively tilted by a few degrees. (S. Ryan (Open University))

Figure 1.20 Cool dense molecular clouds silhouetted against a bright HII region. This particular example is the Horsehead Nebula. (Anglo-Australian Telescope Board)

In the disc, about 50% of the mass of the hydrogen is molecular (H_2). A high molecular content is expected where: (i) the density is high, since this promotes the meeting of atoms; (ii) the temperatures are low, below 100 K, since this avoids the collisional disruption of molecules; and (iii) the UV flux is low, since this avoids the UV-induced disruption of molecules. These are the conditions within the cool dense clouds that are often referred to as molecular clouds (see Figure 1.20). These clouds are found throughout the disc, but are particularly numerous between about 4 and 7 kpc from the Galactic centre. In dense clouds, not only is the hydrogen present mainly as H_2, but similarly all of the chemically reactive elements are predominantly combined into molecules. A particularly important molecule is CO, which is vital for detecting the presence of cold molecular gas (see Box 1.4). There are also some quite large molecules, such as ethanol (CH_3CH_2OH, more often known as 'alcohol'). About 100 different molecules have been detected in dense clouds. In such an environment, only chemically unreactive elements, such as He and Ne, remain predominantly in atomic form.

The relatively small mass of ionized hydrogen (HII) in the ISM is contained in the intercloud media and, much more spectacularly, in HII regions (Figure 1.20). HII regions are frequently associated with dense clouds.

- Gas clouds cooler than about 100 K generally do not emit the 21 cm line; why not?
- At $T < 100$ K, hydrogen forms into H_2 molecules, so no atomic hydrogen remains.

- Why are hot HII regions often found in association with cool, dense clouds?
- New stars form within cool, dense clouds. Only very hot (and therefore massive) stars, particularly the short lived but highly luminous main sequence stars of spectral classes O and B, can ionize hydrogen in their vicinity and thus produce HII regions. As massive stars are short lived, they can be observed still in close association with the original dense clouds. Hence HII regions are found near the cool, dense clouds from which the O and B stars formed.

Interstellar dust

The nature of the soot-like dust grains that form a constituent of the interstellar medium was outlined in Section 1.2.3. The total mass of dust is about 0.1% of the stellar mass of the disc. Dust forms from atoms and molecules in the gaseous ISM that condense directly to solid particles; liquids do not form. The regions of the ISM that favour the formation of dust are those where the density is high and the temperature low. These conditions occur in the disc, where matter ejected from cool giants or supergiants in stellar winds moves away from the star and cools.

BOX 1.4 CARBON MONOXIDE (CO) AS A TRACER OF MOLECULAR GAS

The energy associated with the rotation of a molecule is quantized, analogous to the way that the energy of an electron in an atom is quantized. This means that there are only certain rotational energies that a molecule can have, and changes in the rotational energy of molecules are accompanied by the absorption or emission of a photon. The differences between rotational energy states are very small, and so only low-energy photons are involved; the spectral lines for transitions between two rotational energy states are generally found at radio wavelengths.

However, molecular hydrogen (H_2), and other diatomic molecules that consist of identical atoms (such as C_2, O_2, and N_2) do not emit radiation from rotational energy transitions, for reasons connected with their symmetry. Hence the huge amounts of H_2 in the Galaxy are essentially undetectable. (H_2 does produce some ultraviolet spectral lines, but conditions are seldom favourable for these to be observed.)

Most of the hydrogen in the ISM is in molecular rather than atomic form at temperatures below about 100 K. Consequently, the most abundant element in the Universe effectively becomes invisible at temperatures below 100 K.

However, the carbon monoxide (CO) molecule is composed of two dissimilar atoms, so its rotational transitions *can* be observed. This makes CO, which is reasonably abundant and believed to be distributed in the same way as H_2, an important tracer of cold molecular clouds in space. Carbon monoxide has strong radio emission lines at 1.3 and 2.6 mm that are used for this purpose.

Dust grains emit a continuous spectrum of radiation that is similar to the black-body spectrum. The spectrum of dust may therefore be used to deduce its temperature. If dust particles are heated too much they **sublimate**, that is, they change directly from being solid to being gaseous without melting to form a liquid. The temperature at which sublimation occurs depends on the precise composition of the dust, but even the least volatile compounds (those that are most resistant to evaporation) sublime at temperatures no greater than about 2000 K. In practice though, most interstellar dust grains are well below their sublimation temperature, nearer to 20 K than 2000 K, and are thus easily able to survive in the environment of space. The spectrum emitted by stars peaks in the visible part of the spectrum at wavelengths shorter than 2 μm, whereas the energy emitted by dust grains at a temperature of 20 K peaks around 100 μm (see Figure 1.21). At such a relatively long wavelength, the peak of the dust emission is in the far-infrared region of the spectrum. The distribution of dust can therefore be deduced from far-infrared images, such as Figure 1.14c.

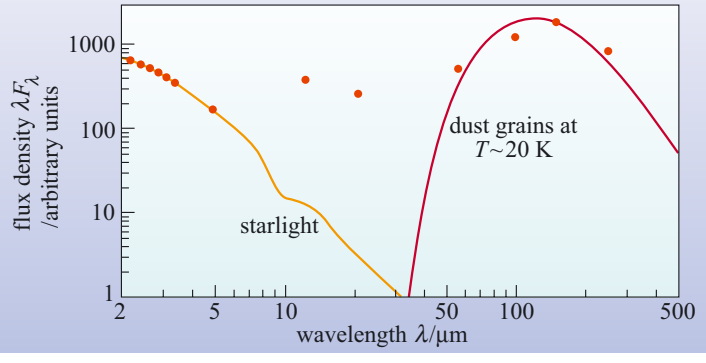


Figure 1.21 The infrared spectrum of radiation emitted by sources in the Galactic plane. The solid dots give the measured spectrum; the curves show the contributions due to starlight, which peaks at $\lambda < 2$ μm (in the visible part of the spectrum), and dust grains at 20 K, which peaks at $\lambda \sim 100$ μm (in the far-infrared). The additional flux around 10 μm to 20 μm comes from large carbon-rich molecules. (Note that this spectrum is shown as $\lambda \times F_\lambda$ against wavelength λ. A full discussion of this type of spectrum is given in Chapter 3.) (Based on data from Li and Draine, 2001)

1.4.3 A cross-section through the disc

In this section we examine the distribution of disc material above and below the Galactic plane. An important aim of this section is to describe the vertical distribution of Galactic components that have no definite boundary. We noted earlier that the

BOX 1.5 THE EXPONENTIAL FUNCTION

The **exponential function** makes use of an important mathematical constant, usually denoted e, which has the value 2.718 to four significant figures. Any mathematical function relates a given value of a variable, x say, to some other value that can be denoted $f(x)$. In the case of the exponential function, this value is given by $f(x) = e^x$ for any value of x. (This last statement is one of many ways of defining the exponential function.) So, if $x = 1$, the exponential function has the value $f(1) = e^1 = 2.718$ (to four significant figures); if $x = 2$ then $f(2) = e^2 = 7.389$; if $x = -1$ then $f(-1) = e^{-1} = 1/e = 0.3679$; and so on.

The exponential function is of great importance throughout physics and astronomy because many natural processes exhibit the phenomenon of *exponential growth* or *exponential decay*, and such processes are described by an equation of the form $y = y_0\, e^{bx}$, where y_0 and b are independent parameters, i.e. two fixed values that may be chosen to fit any particular case. As indicated in Figure 1.22, the chosen value of y_0 will be the value of y when $x = 0$, and the chosen value of b will determine how rapidly y increases or decreases as x changes. Note that positive values of b imply exponential growth, and negative values of b imply exponential decay.

Examples of exponential decay relevant to astronomy include the rate of decay of the radioactive nuclei ^{56}Ni and ^{56}Co which maintain the light output of supernovae, and the absorption of starlight as it passes through an absorbing gas cloud.

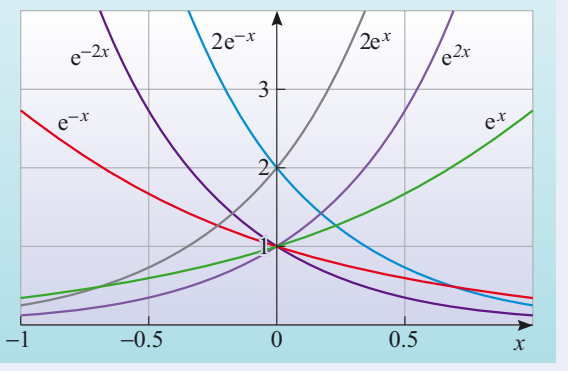


Figure 1.22 Exponential curves described by the equation $y = y_0\, e^{bx}$, for various values of y_0 and b. The value of y_0 is the value of y when $x = 0$, and the value of b determines how rapidly y increases or decreases as x changes.

density of stars declines with increasing distance from the mid-plane of the disc, making it difficult to interpret statements such as: 'the disc has a thickness of about 1 kpc'. Fortunately, the technique introduced here will clarify the meaning of such statements, and thereby allow us to develop a more satisfactory description of the size and structure of the disc. In order to achieve this we make use of a mathematical entity known as the *exponential function* (Box 1.5).

The vertical distribution of stars provides an example of exponential decay. More specifically, the **number density** of stars (i.e. the number of stars per unit volume), n, decreases with distance from the mid-plane, according to the equation:

$$n(z) = n_0\, e^{-|z|/h} \qquad (1.8)$$

Here, the positive quantity $|z|$ represents the distance above or below the mid-plane; h is an important distance parameter called the **scale height** of the disc that characterizes the 'thickness' of the disc; n_0 is the number density of stars in the mid-plane; and $n(z)$ is the number density of stars at a displacement z from the mid-plane, with $z > 0$ for points above the mid-plane and $z < 0$ for points below the mid-plane. The symbol $|z|$ is often read as 'the absolute value of z' or 'the modulus of z', since the modulus sign, $|\ |$, indicates that only the magnitude (as opposed to the sign) of the enclosed quantity should be considered. That is, $|z|$ is always a positive quantity, irrespective of the value of z.

■ What is the ratio of the number density of stars at a distance h from the mid-plane, to the mid-plane number density?

❑ When $z = h$, Equation 1.8 becomes $n(h) = n_0\,e^{-|h|/h} = n_0\,e^{-1}$

You can evaluate e^{-1} using a calculator, or you can note that $e^{-1} = 1/e \approx 1/2.718$. Hence $n(h) \approx n_0 \times 1/2.718 \approx 0.37n_0$, and hence $n(h)/n_0 \approx 0.37$.

That is, at a distance of one scale height above the mid-plane, the number density of stars is a factor 1/e times the mid-plane density, that is, about 0.37 times the mid-plane number density. This illustrates the significance of the scale height, h.

The *scale height* of the disc is the distance over which the number density of disc stars decreases to 1/e times its mid-plane value.

The related ideas of exponential decay and scale height have been introduced in the context of the whole collection of disc stars, but observations indicate that these concepts can, at least approximately, be applied to the various subpopulations of stars in the disc, and even to other entities such as the ISM. Assigning scale heights to subpopulations of stars or classes of objects allows us to describe the vertical structure of the disc far more precisely than could be done by making crude statements about its 'thickness'. The scale height of a subpopulation doesn't say where that subpopulation ends, since there is no 'end', but it does indicate, in a precise way, the distance from the mid-plane at which the density of that subpopulation has significantly decreased. Below, we consider the scale height values for various constituents of the disc, and we comment on the significance of some of those values. We start with the stellar subpopulation belonging to the thin disc that was introduced in Section 1.4.1.

The thin disc

Observations show that the scale heights of stars belonging to the thin disc depend on their spectral classes. That is, thin-disc stars of different spectral types have different distributions above and below the mid-plane of the disc.

Main sequence stars with spectral types G, K or M belonging to the thin disc (including the Sun) are distributed with a scale height of around 300 pc.

■ By what factor does the density of thin-disc G-type main sequence stars change 1.0 kpc from the mid-plane, relative to the density in the mid-plane?

❑ The scale height h for Pop. I G-type main sequence stars is 300 pc. When $z = 1.0$ kpc and $h = 300$ pc, we can write Equation 1.8 as $n(1.0\text{ kpc}) = n_0\,e^{-|1000\text{ pc}|/300\text{ pc}} \approx 0.036n_0$. That is, at 1.0 kpc from the mid-plane, the density of G-type main sequence stars drops to just under 4% of its mid-plane value. This indicates that very few G-type stars of Population I would be found more than 1 kpc from the mid-plane of the disc.

The O and B stars belonging to the thin disc have scale heights of only 50 pc to 60 pc.

Now, we know that O and B stars are very young, whereas the G, K and M stars of the thin disc are Pop. I objects that span the whole age range of that population. It follows that the older stars of the thin disc are likely to be found further from the mid-plane than the more recently formed stars. This observed variation of scale height with age is thought to indicate an evolutionary process that operates within the disc. It is believed most stars form near the mid-plane, but once they have formed they are gradually scattered to greater heights by interactions with giant molecular clouds, which may be as massive as $10^7 M_\odot$.

The thick disc

In Section 1.4.1, it was stated that the stars belonging to the thick disc travel to greater distances from the Galactic plane than the stars of the thin disc. This is reflected in the scale heights associated with this subpopulation.

> The G and K stars belonging to the thick disc have a scale height of approximately 1000 pc to 1300 pc. It is because this scale height is much larger than the scale height of thin-disc stars of similar spectral type (only 300 pc), that the 'thick' disc is so-named.

Thick-disc stars are far less common than thin-disc stars in the mid-plane of the Galaxy. However, the relatively large value of its scale height indicates that the number density of thick-disc stars declines much more slowly than that of thin-disc stars as the displacement from the mid-plane increases. Thus, the thick-disc stars become relatively more important as distance from the mid-plane increases.

It is currently uncertain why the thin disc and the thick disc differ in this way; it could reflect different origins of the two subpopulations, or could be due to some event during the formation of the disc (such as a collision with a dwarf galaxy that added energy to the stars' orbits). This is one issue bearing on the origin and evolution of the Milky Way for which we do not yet have a satisfactory explanation.

The ISM

So far we have compared the scale heights of different classes of stars, but now we consider the vertical distribution of the gas and dust of the ISM. In this case we continue to represent the scale height by h, but rather than consider the exponential decay of a number density of stars, n, we consider the (mass) density ρ of the ISM, measured in kg m^{-3} or some similar unit. Thus, the vertical distribution will be described by an equation of the kind $\rho(z) = \rho_0 \, \mathrm{e}^{-|z|/h}$.

> The ISM is more concentrated towards the Galactic plane than most G and K stars; it has a scale height around 150 pc.

The fact that the scale height of young O and B stars, 50 pc to 60 pc, is *less* than that of the gas from which they formed tells us about the star formation process. The rate at which stars are formed from the ISM is called the **star formation rate**, **SFR**, and may be measured in solar masses per year in any specified region (often the whole Galaxy). If the star formation rate were proportional to the local density of gas, that is, if SFR $\propto \rho$, then stars would be formed with the same height distribution as the gas. However, the O and B stars are *more* concentrated towards the mid-plane than is the gas. This indicates that the simple proportionality between

SFR and ρ cannot be correct. Since young stars are more concentrated towards the mid-plane than is the ISM, it must be the case that star formation is more effective when the ISM density is higher. (This implies that any simple 'power law' relating SFR to ρ must take the form SFR $\propto \rho^n$, where n is a number greater than 1.) We have thus used the vertical distribution of gas and young stars in the disc to infer something about how the star formation rate depends on the density of gas.

■ Can you state the scale height for the Sun? Justify your answer.

❑ The concept of a scale height has no meaning for a *single* object; it characterizes how the density of a *class* of objects varies, so it cannot be applied to individual members of the class.

1.4.4 The spiral arms

So far in this section we have examined the constituents of the disc, their distribution and composition. We have also seen that the Milky Way has a thin stellar disc, and in this way resembles the spiral galaxies we observe beyond our Galaxy. So it is not surprising that we find evidence for spiral structure in the Milky Way. In this section we look at the structure of the spiral arms and their possible origin.

Tracing spiral structure

As indicated in Section 1.2.1, it is believed that spiral arms stand out mainly because they contain concentrations of *bright* objects associated with recent star formation rather than being strong concentrations of *mass*. O and B stars are so short lived that, when you see them, they cannot be far from where they were born, so it is usual to regard the spiral arms as the main locations of star formation in the Milky Way.

At optical wavelengths, dust makes it difficult to see stars in the disc of the Milky Way if they are more than a few kiloparsecs away. Consequently, the spiral structure of the Galaxy is more easily mapped by using a combination of optical and radio 21 cm observations, since the latter are unaffected by dust and permit the detection of neutral hydrogen to much greater distances. A range of objects has been observed that appear to map out spiral structure. Such objects include dense molecular clouds, HII regions, open clusters, and OB associations. Objects which are used to map the locations of the spiral arms are called **spiral-arm tracers**.

The map shown in Figure 1.23a (overleaf) indicates the distribution of bright HII regions, prominent clusters of young stars, and dense clouds. They seem to trace out three parallel strips near the Sun that can be interpreted as neighbouring spiral arms. An artist's conceptualization of this and other data (Figure 1.23b) shows the location of the Sun relative to the Galactic centre and the local spiral arms, named the **Sagittarius–Carina Arm**, the **Orion–Cygnus Arm** and the **Perseus Arm**. The Sun seems to be contained within the Orion arm, although it is unclear whether this is really a spiral arm in its own right or simply a 'side spur' belonging to some other arm; sometimes it is referred to as the **Orion Spur**. When looking at artists' conceptions of the Milky Way (e.g. Figure 1.23b) that include beautifully unbroken spiral arms, remember that scientific backing for such detailed views is almost entirely lacking. As noted in the introduction to this chapter, there is also evidence of a central bar in the Milky Way, which is rarely shown in older artistic pictures.

The winding dilemma

We have discussed how the spiral arms can be traced using objects associated with recent star formation. However, in the remaining parts of this section, you will see that it is difficult to explain *why* star formation occurs along these reasonably well-defined spiral tracks. It is also difficult to explain their persistence over long periods of time.

■ How do we know that spiral arms last for long periods of time?

❑ Observations of the Milky Way do not tell us this. However, astronomers see many other spiral galaxies and are hence able to infer that spiral arms are not just transient (short-lived) structures.

In trying to account for the existence and persistence of spiral arms, one point must be kept clearly in mind: if the arms were composed of an *unchanging* set of stars, the differential rotation of the disc would cause the shape of the arms to alter with time, so an initially 'realistic' pattern of arms would soon cease to resemble any observed spiral arms. This is called the **winding dilemma**, and is explored further in Question 1.9.

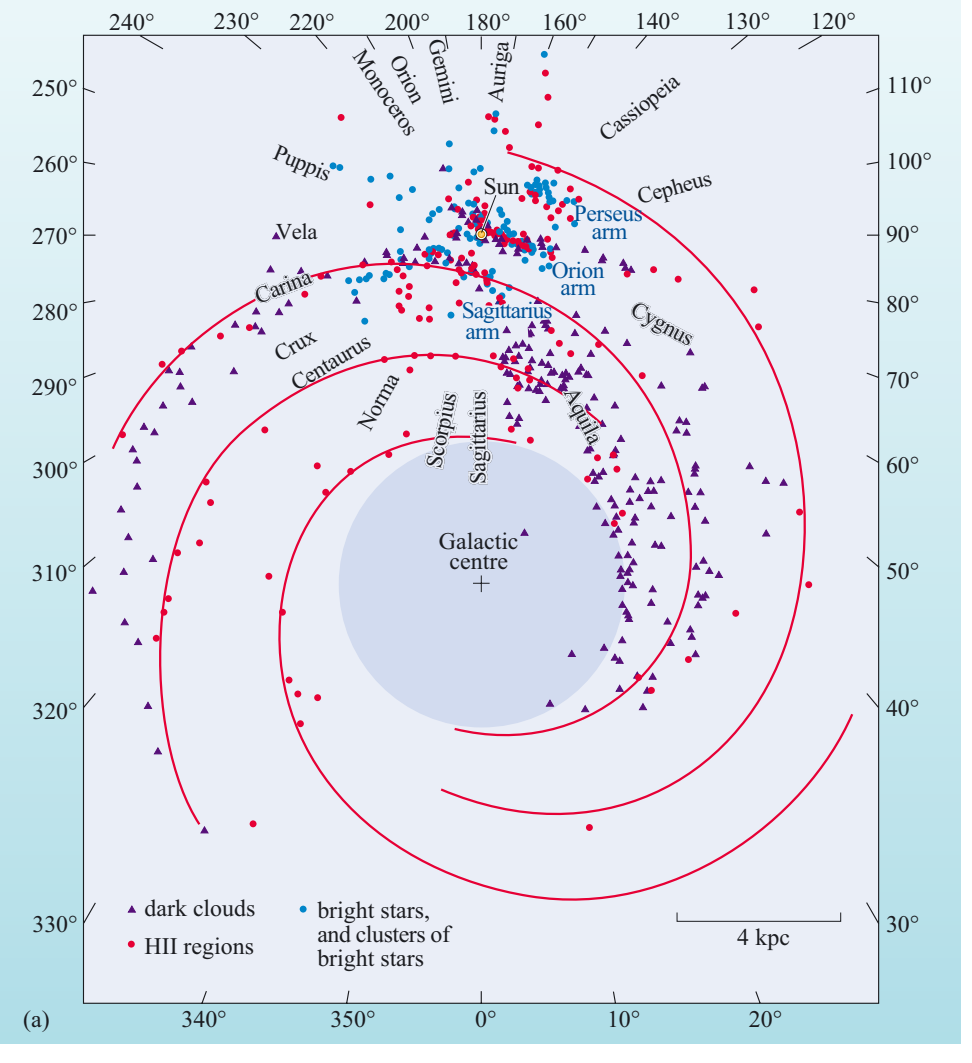


Figure 1.23 (a) The spiral arms of the Galaxy as mapped out by the distribution of dark clouds, HII regions and prominent clusters of young stars. This is a face-on view of the Galactic disc, with galactic longitude shown around the perimeter (note that the galactic coordinate system is centred on the Sun and *not* on the Galactic centre). (Composite of data from Georgelin and Georgelin, 1971, 1976; Grabelsky *et al*., 1988; and Binney and Merrifield, 1998, from data kindly provided by P. Solomon)

QUESTION 1.9

Imagine that all disc stars have the same rotation speed of 220 km s^{-1} for all radial distances from 4 to 10 kpc from the Galactic centre, and suppose that a pattern of spiral arms like that shown in Figure 1.23b already existed in the Galaxy 4.5×10^9 years ago, when the Sun formed.

(a) How many orbits of the Galactic centre would the inner end of one of the arms, located 4 kpc from the Galactic centre, have completed since the formation of the Sun?

(b) How many orbits would the outer end of one of the arms, located 10 kpc from the Galactic centre, have completed in the same time?

(c) If the spiral arms had been made of the same stars throughout the lifetime of the Sun, what would they look like now? How well does your answer correspond to the actual appearance of spiral arms?

Even allowing for its oversimplifications, the answer to Question 1.9 indicates that there is no permanent population of 'spiral arm' stars. This suggests that the bulk of the stars that are currently in spiral arms must be moving relative to those arms, and will not remain in them for long.

Although spiral arms may represent a persistent pattern, they cannot always be made of the same stars throughout their lifetime.

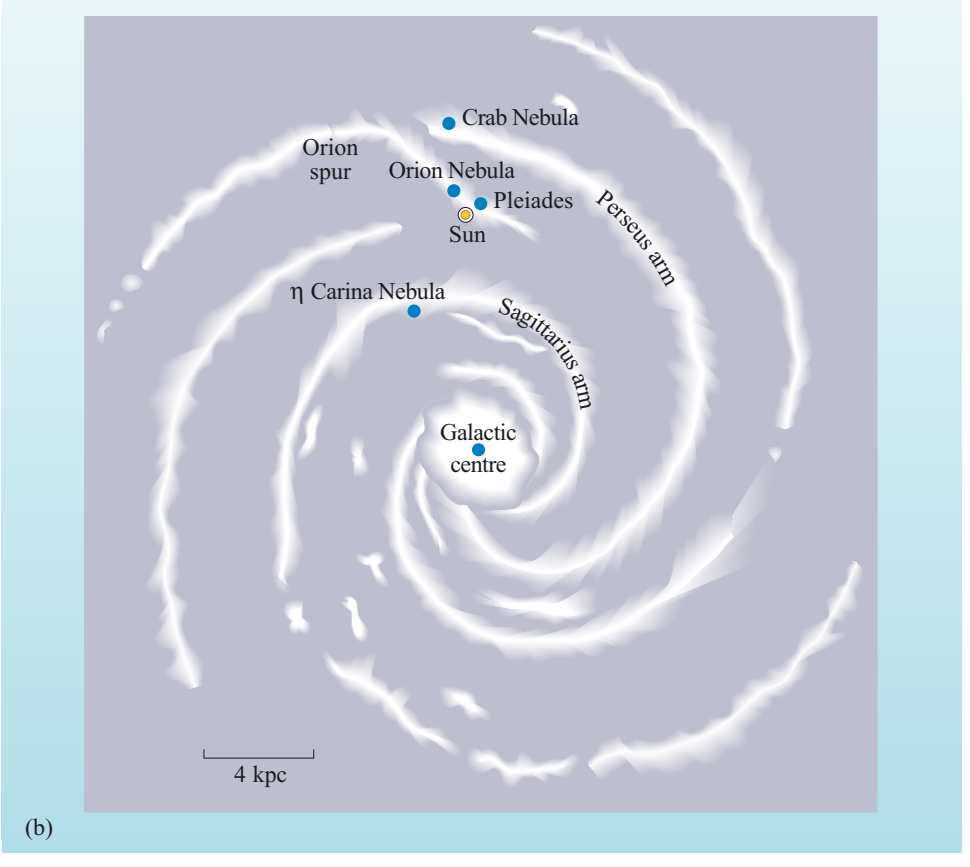


Figure 1.23 (b) Artist's schematic of the location of the Sun relative to the spiral arms. The bar/bulge is not shown in this highly schematic view.

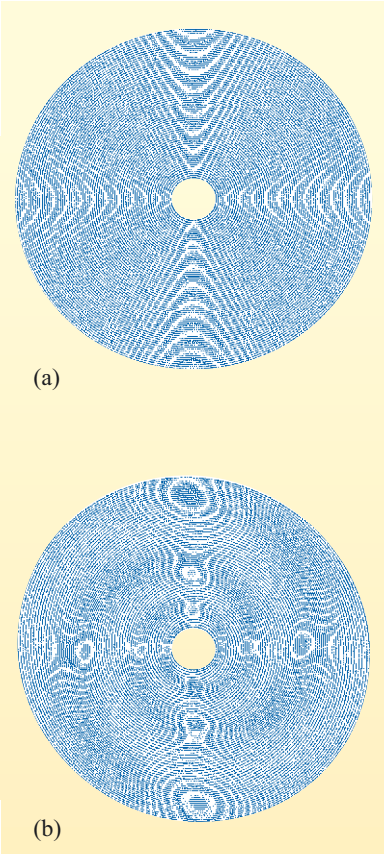


Figure 1.24 (a) Nested elliptical orbits with the long axes of the ellipses aligned in the same direction. (b) The same orbits as in (a), but with each ellipse rotated slightly relative to the adjacent one, giving rise to two spiral density waves, through which orbiting material can pass. (S. Ryan (Open University))

How is the persistence of spiral arms reconciled with the differential motion of the stars and clouds that trace them? Various answers have been proposed to this question. One idea is that the spiral arms represent the current location of 'waves' of star formation that are travelling through the disc. Such waves need not participate in the differential rotation of the disc any more than sound waves travelling through air have to travel at the speed of the wind. If the waves formed a spiral pattern, as might be natural in a differentially rotating disc, then a persistent pattern of spiral arms might be an expected feature of the Galaxy. Below we outline one specific proposal regarding the origin of such waves of star formation.

Density wave theory

The **density wave theory** was developed by C. C. Lin and Frank H. Shu in the 1960s. They treated the disc as an approximately smooth, axially symmetric distribution of matter in a state of steady differential rotation, but they assumed that such a disc would naturally develop regions in which the density was enhanced relative to the surrounding material. Lin and Shu argued that certain spiral patterns of density enhancement were especially favoured and could become self-perpetuating. Long-lived patterns of density enhancement of this kind are called **spiral density waves**.

Figure 1.24 provides a highly simplified view of a spiral density wave. A set of ellipses is shown in Figure 1.24a, representing orbiting stars and gas in the disc of the Galaxy. The long axes of all orbits in Figure 1.24a point in the same direction. However, if each successive orbit is rotated slightly, as in Figure 1.24b, then a spiral pattern emerges. If such a pattern of orbits could persist, then the disc would show a permanent pattern of density enhancements, even though the material responsible for that pattern would change continuously as orbiting material moved through the regions where the arms are seen. In practice, such a simple pattern of orbits would not persist, due to the gravitational interaction of material in neighbouring orbits. The challenge addressed by Lin and Shu was to show that some such pattern of spiral density enhancements could arise and persist in a more realistic differentially rotating disc.

A simple spiral density wave is expected to rotate about the Galactic centre, but to maintain its shape as it does so. That is, the pattern rotates *rigidly*, despite the fact that the material from which it is made rotates *differentially*. In fact, across most of the disc, the density wave moves more slowly than the matter in the disc. Only towards the outer edge of the pattern does the rotation speed of the density wave equal that of the disc. The radius at which this occurs is called the **co-rotation radius**. This means that throughout most of the disc, stars and gas approach the slowly moving density wave from behind, pass through it, and then move ahead of the wave's leading edge. Gas is compressed when it enters the density wave. For a giant molecular cloud on the verge of forming stars, this increase in density might be enough to trigger star formation and thus account for the presence of the young objects that trace the arm.

Despite the reasonableness of these ideas, density-wave theory could hardly be said to be well-confirmed. In particular, there is substantial doubt about the mechanism that produces and (at least temporarily) maintains the density wave. So, despite the great importance of spiral arms in the Milky Way's disc, their origin should still be regarded as uncertain at present.

QUESTION 1.10

Assuming that the co-rotation radius of the Milky Way is at 15 kpc from the Galactic centre, and using the Galactic rotation curve given in Figure 1.13c, draw the rotation curve of a spiral density wave and estimate the speed at which the Sun would approach such a wave.

1.5 The stellar halo and bulge of the Milky Way

We now turn our attention from the disc – which is a region of ongoing star formation – to the stellar halo and the bulge. These two regions, which together form the Galactic spheroid, are locations where star formation has long since ceased, and where the oldest stars in the Galaxy are found. It is by studying these regions that we might hope to find clues about the early evolution of the Galaxy. With this in mind, we now proceed to examine these components of our Galaxy.

The stellar halo consists mainly of very old Population II stars which have low metallicities and which move in roughly elliptical orbits (Figure 1.12) that are often highly inclined to the Galactic plane. Stars following such orbits plunge through the disc from time to time, but the spaces between stars, even in the disc, are so great that collisions are highly improbable.

■ Which spectral classes of main sequence star would you expect to be common in the stellar halo? What other kinds of star are likely to be abundant there?

❑ Only long-lived stars should be common. Apart from old main sequence stars of spectral classes K and M, there should be other, more highly evolved, stars such as red giants, horizontal branch stars, and stars belonging to the asymptotic giant branch of the H–R diagram. Even more highly evolved objects, such as white dwarfs, might also be expected.

Measurements indicate that the stellar halo is somewhat flatter than a perfect sphere. Shape determinations are based on counts of the number of halo stars in a plane surrounding the Sun that is perpendicular to the direction of the Galactic centre (see Figure 1.25). Stars in this particular plane at a given distance from the Sun will all be at a common distance from the Galactic centre, so counts of such stars can be used to assess the shape of a section through the stellar halo. Such measurements indicate that the ratio of the axes a and c, shown Figure 1.25, is $c/a \approx 0.8$, making the stellar halo an oblate spheroid.

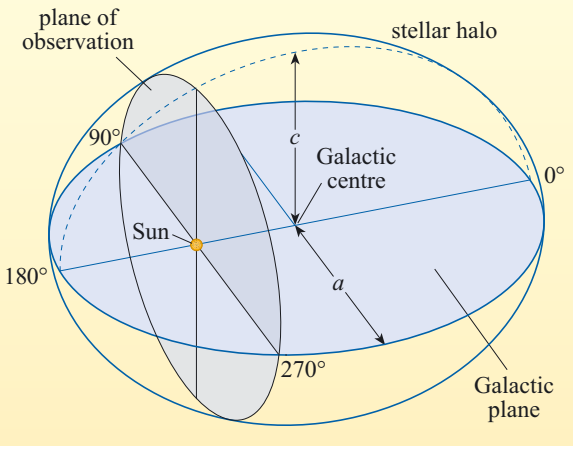


Figure 1.25 The shape of the stellar halo is determined by counting stars in a plane of observation cutting vertically through the Galaxy at the location of the Sun. Counts made at various Galactic latitudes determine the ratio of axes c/a (or axial ratio) of the stellar halo.

1.5.1 Globular clusters

Globular clusters were introduced in Section 1.2.4, where it was stated that about 1% of halo stars are found in these objects and that they typically contain 10^4 to 10^6 stars in a spherical region up to about 50 pc across. They are relatively prominent, and can easily be picked out and studied in unobscured regions of the sky. There are probably between 150 and 200 globular clusters in the Milky Way, about two-thirds of which are associated with the halo.

The stars in a globular cluster are most densely packed toward the centre. Because of this central concentration, the central regions of globular clusters are often overexposed in photographs. The number density of stars at the *centre* of a globular cluster is typically $10^4\,\mathrm{pc}^{-3}$. This is approximately 10^5 times higher than the number density in the solar neighbourhood (~$0.1\,\mathrm{pc}^{-3}$), but it still leaves plenty of space between individual stars.

QUESTION 1.11

(a) What is the average separation of stars in the centre of a globular cluster? Give your answer in pc. (*Hint*: consider the average volume occupied by a single star.)

(b) How does this compare to the distance between the Sun and the stellar system of alpha Centauri, which is 1.31pc away?

The distribution of globular clusters

Globular clusters played a key role in developing our view of the size and shape of the Galaxy. Globular clusters are not distributed uniformly around the sky, but are concentrated in the direction of the constellation of Sagittarius. This was first recognized by Harlow Shapley, and is shown quite clearly in Figure 1.26a, which is a map of the positions of halo globular clusters on the sky. (This complements the side-on view of Figure 1.10.) In 1917, on the basis of his studies of globular clusters, Shapley asserted that the centre of the Milky Way was in the direction of the constellation of Sagittarius. Realizing that globular clusters were more numerous near the Galactic centre, he correctly reasoned that the centre of their distribution probably indicated the centre of the Galaxy.

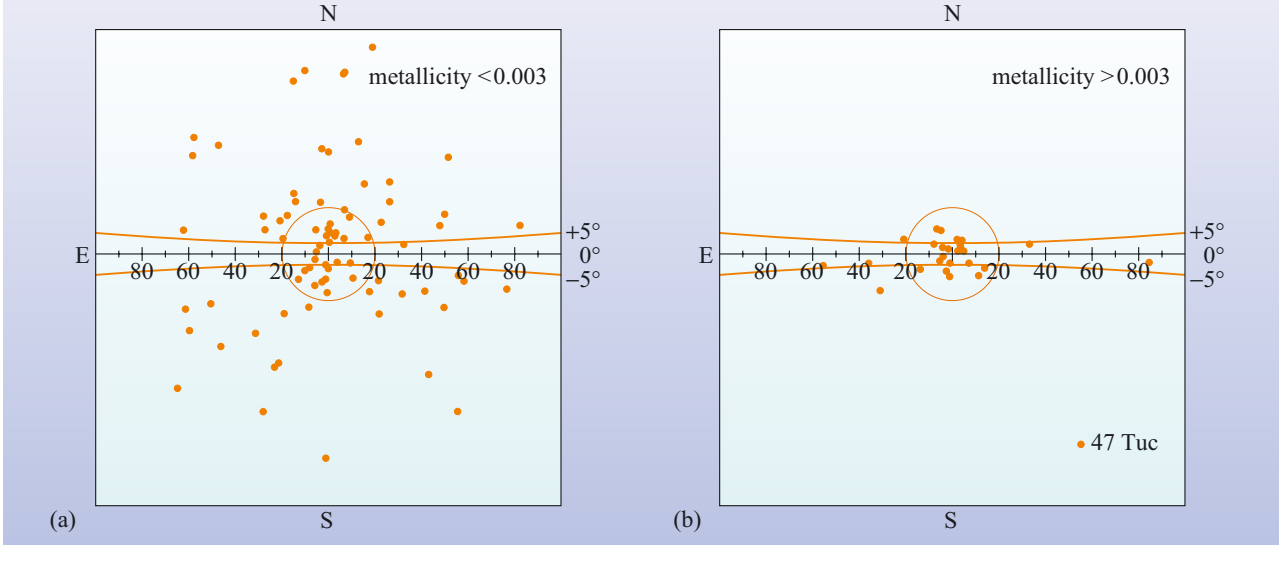


Figure 1.26 The locations of globular clusters on maps of the sky centred on the Galactic centre. The horizontal line is the Galactic plane, and the shallow curves mark Galactic latitudes ±5°. The circle has a radius of 20°, which corresponds to ≈3 kpc at the distance of the Galactic centre where the globular cluster distribution is centred. Map (a) shows globular clusters with very low metallicity $Z < 0.003$ (the halo globular clusters), while map (b) shows clusters with $Z > 0.003$ (the disc globular clusters). The latter are much more concentrated towards the Galactic centre and the Galactic plane. (Zinn, 1985)

HARLOW SHAPLEY (1885–1972)

Harlow Shapley (Figure 1.27), the son of a farmer, was born in Missouri, USA. After a limited education he began work as a crime reporter on a small Kansas newspaper. Further study led to the University of Missouri, where he planned to take a degree in journalism. However, the School of Journalism had not yet opened, so rather than waste a year he took an astronomy course and thereby began one of the most distinguished astronomical careers of the 20th century. He gained his PhD from Princeton in 1913, was a staff member at the Mount Wilson Observatory until 1921, and was then appointed Director of the Harvard College Observatory where he remained until 1952. Shapley began his studies of Cepheid variables in globular clusters while at Mount Wilson. Using the Cepheids as distance indicators (see Chapter 2) he found that the known globular clusters were concentrated about a location in the direction of Sagittarius that he identified as the centre of the Milky Way. Shapley overestimated the diameter of the Milky Way's disc by a factor of three, but he did comprehend its structure, and he recognized the off-centre location of the Sun. In a long and illustrious career, Shapley made many contributions to the study of the Milky Way and other galaxies.

Figure 1.27 Harlow Shapley. (Bachrach Portrait Studios)

Because globular clusters are so conspicuous, and hence can be recognized at great distances, they provide a potential means of determining the size of the stellar halo. Their distances can be derived from the brightness of their stars. Observations show that although there are many globular clusters within 20 kpc of the centre of the Galaxy, and a few beyond 37 kpc, there are none in between. Some astronomers treat the break at 20 kpc as a rough indication of the outer edge of the stellar halo, while others regard the more distant ones as indicating just how very extensive the stellar halo is! Unfortunately, the low number of globular clusters in the Galaxy – there are probably no more than about 200 – prevents any improvement in estimates based on globular clusters alone. In order to probe the outer part of the stellar halo, it is necessary to use the more numerous, but harder to recognize, non-cluster stars. Such objects are discussed in the next section.

Not all the globular clusters actually belong to the halo. Globular clusters with metallicity $Z > 0.003$ used to be regarded as halo objects, but are now recognized as belonging to the thick disc. They account for approximately one-third of all globular clusters in the Galaxy. Figure 1.26b shows the location of these relatively high-metallicity clusters. As you can see, they are mostly confined to the Galactic plane, as is appropriate for thick-disc objects.

Globular cluster ages

The halo stars are the oldest known stars. Those that are in globular clusters are particularly important because of the relative ease with which their ages can be determined. The globular clusters of the halo formed early in the evolution of the Milky Way, probably even before the Galaxy was a well-defined entity, so they can teach us something about the formation of the Milky Way. In addition, these oldest globular clusters provide a lower limit for the age of the Universe. Obviously, the Universe must be older than the oldest globular clusters it contains. There have been

times when the observationally determined globular cluster ages have exceeded the age of the Universe estimated from some cosmological arguments. Such conflicts have been a source of great controversy, and their resolution is a significant sign of progress.

The age of a globular cluster is determined by analysing its Hertzsprung–Russell or H–R diagram. In order to see how this is done, we need to consider how the H–R diagram is drawn. Box 1.6 provides a description of the different forms of the H–R diagram that we shall refer to.

BOX 1.6 H–R DIAGRAMS

Theoretical models of stars generally provide calculated surface temperatures and luminosities. The theoretical **H–R diagram** of a set of stars is therefore a plot of luminosity versus (surface) temperature. Such a plot is sometimes called a *temperature–luminosity diagram*. However, the temperatures and luminosities of real stars cannot be measured directly; they must be inferred from quantities that can be measured. It is quite common to plot the H–R diagram of real stars using the observed properties that correspond most closely to temperature and luminosity, which are *colour* and *brightness*.

■ Why does colour correspond to temperature?

❑ Stars behave similarly to black bodies, and the continuous spectra emitted by black bodies peak at shorter wavelengths the hotter they are, so hotter stars look bluer.

Instead of just saying that a star is blue or red, astronomers express colour in a quantitative way by comparing the amount of energy received in one part of the spectrum with the amount of energy in another. One common way of measuring colour is to compare the energy received in the blue (B) part of the spectrum with the energy received in the green part. Since the green part of the spectrum is where the human eye has maximum sensitivity, this part of the spectrum is called the visual (V) band. From the ratio of the energy between the blue and visual bands, astronomers define a value they call the **colour index**, $B - V$, (pronounced 'B minus V'). For hot, white stars, $B - V \sim 0.0$, whereas for cool, red stars, $B - V \sim 1.0$. (You need not be concerned here about the details of how exactly the $B - V$ value is calculated from the observations.)

The brightness of a star is usually expressed using a magnitude scale, which describes its brightness relative to other stars. A difference of 2.5 magnitudes between two stars corresponds to a factor of 10 in brightness. (It could be considered as a *faintness* scale, as stars whose magnitudes are at the positive end of the magnitude range are the faintest ones, while the brightest stars are at the negative end!) Magnitude determinations are often restricted to a specified part of the spectrum, such as the visual (V) band. Within such a specified band, a star has two magnitudes that are of interest. The first of these is the easily determined **apparent visual magnitude** (m_V), which directly compares the apparent brightness of stars, even though those stars may be at very different distances. The other kind of magnitude is the **absolute visual magnitude** (M_V), which compares the brightness that the stars would have if they were all at the same distance. It is this latter quantity that is most directly related to their intrinsic luminosity, but it is also the harder to determine accurately. The two kinds of magnitude are defined in such a way that their values would be equal for a star that was 10 pc away.

The precise relationship between the apparent and absolute visual magnitudes of a star at a distance d is given by

$$M_V = m_V - 5 \log_{10} (d/\text{pc}) + 5 \qquad (1.9)$$

where the distance is expressed in parsecs.

In contrast to the theoretical temperature–luminosity diagram, the observational version of the H–R diagram plots a colour index on the horizontal axis and a magnitude scale on the vertical axis. Sometimes a diagram of this type is called a *colour–magnitude diagram*. Stars map out similar patterns in both the theoretical and observed versions of the diagram, as the physical information captured in the H–R diagram is essentially the same even in these different forms, and we use the name H–R diagram to refer to all the different forms.

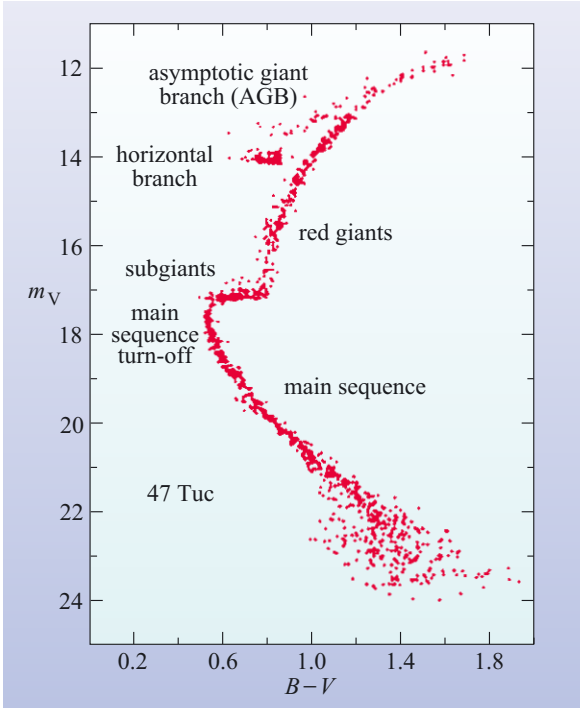


Figure 1.28 The H–R diagram of the globular cluster 47 Tuc. Some of the major features of the H–R diagram have been labelled. (Hesser *et al*., 1987)

Figure 1.28 shows the H–R diagram of the globular cluster 47 Tuc, which is broadly representative of most globular clusters. It is presumed that the stars comprising a cluster all formed at the same time and had the same composition. When it was still young, all the cluster's stars would have belonged to the main sequence and the cluster's H–R diagram would have looked something like that of NGC 2362 in Figure 1.29a. As the cluster aged, the more massive stars of high luminosity would have left the main sequence and evolved to become supergiants. The point on the cluster's H–R diagram corresponding to stars that are just reaching the end of their time on the main sequence is called the **main sequence turn-off**. This is clearly seen in the cluster H–R diagram of Praesepe (Figure 1.29b) – the main sequence turn-off in this case is at $B - V \approx 0.2$, $M_V \approx 1$. With the further passage of time, the main sequence would have become progressively depopulated of massive stars. The H–R diagram of a middle-aged cluster would have looked something like that of M67 (Figure 1.29c) – which begins to resemble the old-cluster diagram, Figure 1.28. The age of an individual star cluster can be gauged by making a calculation of the time required for the H–R diagram to evolve into the observed form.

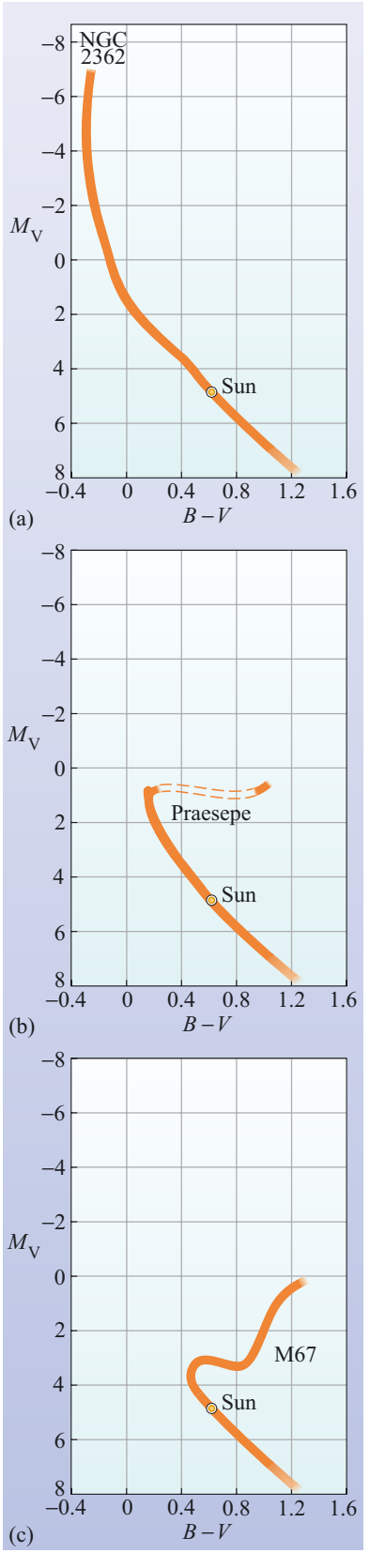


Figure 1.29 The H–R diagrams of three clusters of stars. (a) When the cluster is only a few million years old, as in the case of NGC 2362, essentially all its stars belong to the main sequence. (b) After a hundred million years, the age of the Praesepe cluster, all the high-mass stars will have left the main sequence. (c) After a few billion years, the age of M67, only low-mass main sequence stars remain. The surviving intermediate-mass stars will have evolved into red giants. All of these clusters are open clusters. As in the case of globular clusters, it is believed that all the stars in a given open cluster were formed at the same time from material of uniform composition. (Based on Arp, 1958)

The process of making a theoretical estimate of the time required for a cluster's H–R diagram to acquire a specific form involves the use of computer programs to model the evolution of the stars that make up the observed H–R diagram. Theoretical stellar models are much simpler than real stars, so they must be made as realistic as possible. Some aspects of stellar modelling are particularly difficult, notably the treatment of convection. Despite the difficulties, a good deal of effort has gone into the computation of theoretical cluster H–R diagrams. As a result of such calculations, it is possible to plot the theoretical positions on an H–R diagram of stars in a cluster for a particular age of that cluster. Such a plot defines a curve on the H–R diagram that is called an **isochrone**. The most age-sensitive feature of a set of isochrones of differing ages is the location of the main sequence turn-off, so it is this that is used to date the clusters. Figure 1.30 shows a comparison between the predicted forms of the turn-off and the observations for one particular cluster (M92). As you can see, the age indicated in this case is about 16×10^9 years, although there is clearly some uncertainty, more so in the models than in the observations. Improved calculations show that globular cluster ages are in the range (12 to 15) $\times 10^9$ yr.

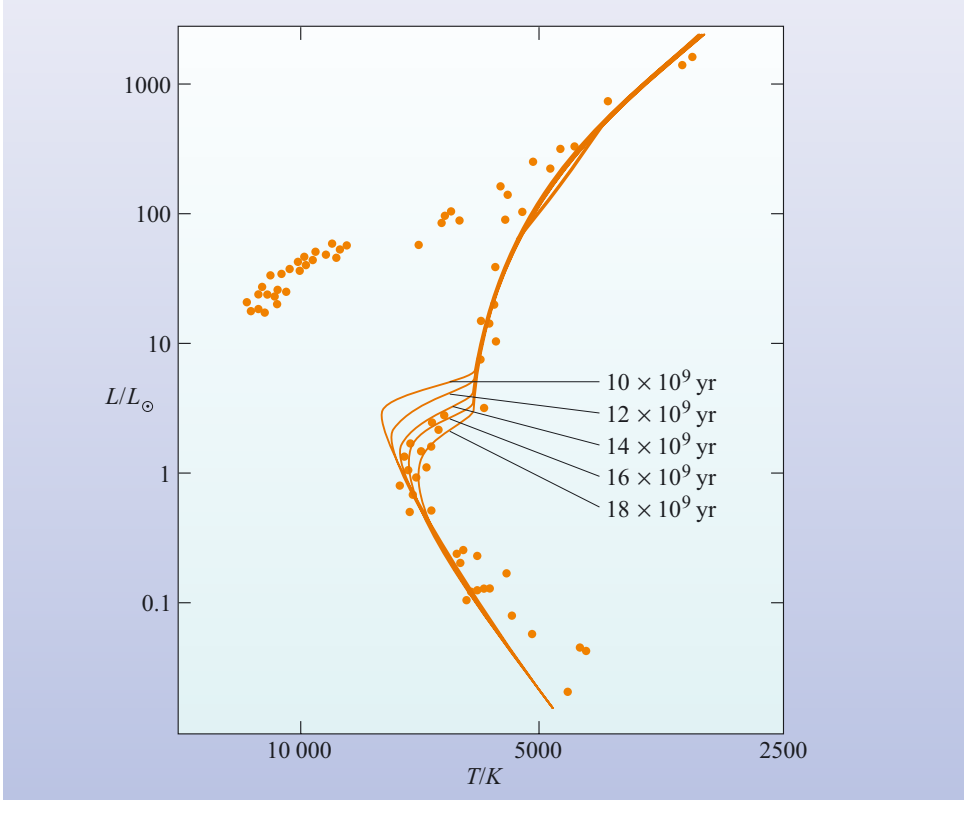


Figure 1.30 Theoretical isochrones showing predictions of the main sequence turn-off points in a globular cluster at various ages, compared with observational data for the globular cluster M92. (Mihalas and Binney, 1981)

Because of difficulties in finding precise *absolute* ages, there has been more progress with the question of the *relative* ages of different globular clusters. Studies indicate that some of the halo clusters are several billion years older than others. If this is correct, it indicates that the formation of the stellar halo was not, as was once thought, the result of the rapid collapse of a galaxy-sized gas cloud over just a few tens of million years. Either the collapse was a much more gradual process, or the stellar halo didn't form in a collapse at all, but rather by the coalescence of many different clouds. Under this second scheme, globular clusters formed either in these clouds or in collisions between them. In either case, the range of globular cluster ages indicates the timescale of the halo formation process.

1.5.2 RR Lyrae stars

The small number of globular clusters limits their usefulness as probes of the vast stellar halo. Fortunately, non-cluster stars belonging to the halo greatly outnumber globular clusters, and these stars can be used to probe the halo further. Members of a particular class of variable stars, known as **RR Lyrae stars**, are especially valuable for this purpose. These stars are named after the first star of their class to be studied, the star designated 'RR' in the constellation Lyra.

Low-metallicity, low-mass stars that have already been through the main sequence and red giant stages of evolution, and have become hot enough in their cores to burn helium, are called '**horizontal branch**' stars (see Figure 1.28). This name was given because all such stars have similar luminosities irrespective of their effective temperature, and they therefore form an almost horizontal band in the H–R diagram. RR Lyrae stars are a subset of the horizontal branch stars, and therefore have known absolute magnitudes.

Objects that, as a class, have known absolute magnitudes, are called '**standard candles**' because their known 'standard' light output allows the distance of any member of the class to be determined from a measurement of its apparent magnitude. (Uses for other standard candles in gauging the distances to other galaxies are discussed in Chapter 2.) Horizontal branch stars meet this criterion, which makes RR Lyrae stars excellent standard candles, *provided* they can be recognized. How do you recognize an RR Lyrae star? They are the subset of horizontal branch stars falling in a region of the H–R diagram within which the outer layers of stars are unstable and pulsate. This region of the H–R diagram is called the **instability strip**. It is the pulsational properties of RR Lyrae variables that make them recognizable.

- Horizontal branch stars that do not lie within the instability strip also exist. Could these be used easily as standard candles too?
- The absolute magnitudes of non-variable horizontal branch stars are known, so they could in principle be used as standard candles. However, there is a practical difficulty: when you observe a non-variable star, it is not usually easy to tell whether it lies on the horizontal branch unless you already know its distance, in which case you don't need a standard candle. It is the distinctive variability of the RR Lyraes which identifies beyond doubt that they lie on the horizontal branch.

The known absolute magnitudes of RR Lyraes, and the relative ease with which they can be recognized, makes them useful for studying the distribution of matter in the halo.

QUESTION 1.12

The equation which relates the absolute and apparent magnitudes of an object to its distance d (in parsecs) (Equation 1.9) can be expressed as

$$m_V - M_V = 5 \log_{10}(d/\text{pc}) - 5$$

The absolute magnitudes of RR Lyrae variables are $M_V \approx +0.5$. Assuming a well-equipped telescope could detect stars down to an apparent magnitude $m_V \approx +20.5$, to what distance could they be seen?

As RR Lyrae stars can be seen out to distances of tens of kiloparsecs, they are very useful probes of the stellar halo. They show that the number density of stars in the halo (i.e. the number per unit volume) falls off with distance from the Galactic centre, r, roughly in proportion to $1/r^3$, at least out to about 30 kpc. This is compatible with the decrease in number density seen for the globular clusters. As the globular clusters and RR Lyrae stars share the same spatial distribution, age, and metal content, it is fair to assume they are part of the same population.

We saw above that globular clusters provided the basis of one of the first reliable measurements of the direction and distance to the Galactic centre. RR Lyrae stars are also valuable for this purpose. Astronomers can observe RR Lyrae stars in the direction of the Galactic centre, and calculate the distance of each one they see. By noting the distance at which the number density of RR Lyrae stars reaches a maximum, astronomers have inferred a distance to the Galactic centre of 8.7 ± 0.6 kpc. This is clearly consistent with the value 8.5 kpc that is often adopted.

We have seen previously that cosmic recycling leads to an increase in the metallicity of the Galaxy with time. Just as stars convert light elements into metals, they also convert hydrogen into helium, the second most abundant element in the Universe. So, due to cosmic recycling, there should be some change in the abundance of helium over time, and the old Pop. II stars of the halo might be expected to contain a smaller proportion of helium than younger Pop. I stars such as the Sun. The helium content of stars is generally very difficult to measure, but fortunately the pulsational properties of RR Lyraes depend on their helium content. Thanks to this, RR Lyrae stars made possible one of the earliest measurements of the helium content of Pop. II stars. In halo stars, the helium fraction in the stellar envelope appears to be around 0.24 to 0.25 by mass, meaning that a quarter of the mass of the stellar envelope is helium, whereas in the Sun it is slightly higher at 0.27 to 0.28. This provides evidence of how much star formation has occurred over the time between the formation of halo stars and the formation of the Sun.

1.5.3 The Galactic bulge

We have seen how the globular clusters and RR Lyrae stars occupy the relatively sparsely populated spheroid, and have provided much information about the structure, size and age of the Galaxy, but there is an equally old component in the Galaxy that has a much greater density. This is the bulge, which occupies the central few kiloparsecs of the Galaxy. In the remainder of this section we investigate the structure of the bulge and the very centre of the Galaxy.

Our view of the bulge is heavily obscured by dust, particularly the central region, which is the meeting place of the densest parts of the halo and the disc. However, it can be observed at infrared wavelengths (this was shown in Figure 1.14b) where it manifests itself as a concentration of brightness, and a thickening around the Galactic centre. The equatorial radius of the bulge is about 3 kpc.

While the majority of bulge stars are probably as old as the stellar halo, many of them have the same metallicity as the Sun. This combination of great age and high metallicity suggests that star formation proceeded very rapidly in the bulge, so that a large degree of cosmic recycling was achieved very quickly. This is consistent with the relatively high density of the bulge compared with the halo, since the star-formation rate, and hence the metal-enrichment rate, would be expected to be greater in a region of greater density.

Unlike the situation in the disc, where stars travel in almost circular orbits around the Galactic centre, the motions of stars in the bulge appear to be more diverse, with speeds in its outer regions around 100 km s^{-1}. Studies of the bulges of other galaxies, combined with available information about the Milky Way, have assessed whether the bulges of galaxies are oblate spheroids that are symmetric about the galaxy's rotation axis, or, alternatively, whether they have three axes of different length. For the Milky Way, the evidence points to the bulge being a triaxial bar. The first suggestion that the Milky Way might have a central bar was made by Gerard de Vaucouleurs (1918–1995) in 1964, based on the motion of HI gas in the inner regions of the Galaxy. The studies described below make use of more recent evidence.

The distribution of diffuse starlight

Infrared observations at a wavelength of 2.4 μm reveal diffuse starlight from an old population of stars towards the Galactic centre. They show an enhancement to one side of the centre, suggesting an asymmetry in the distribution of stars around the central region. This asymmetry can be interpreted as evidence of a bar-like distribution of stars, as indicated by the innermost central contour in Figure 1.31. Its discoverers associate this small bar with the central feature seen in the infrared image of the Galaxy (Figure 1.14b), which was obtained using the Diffuse Infrared Background Experiment (DIRBE) on the Cosmic Background Explorer (COBE) satellite.

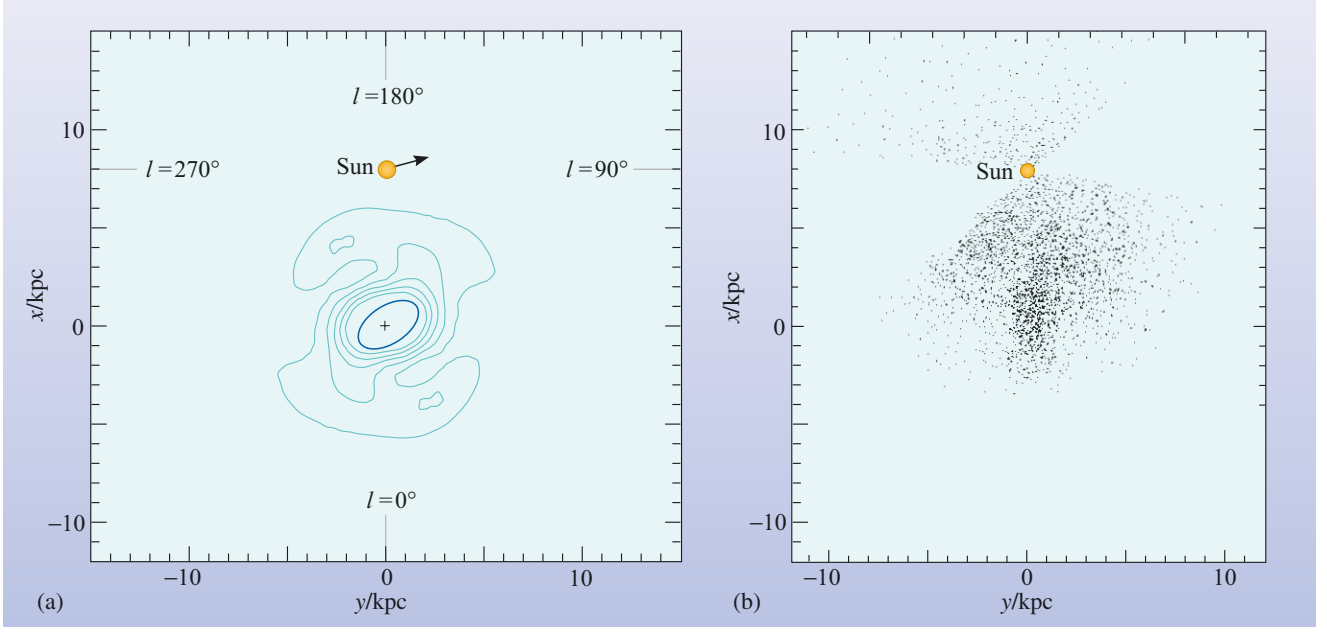


Figure 1.31 (a) Schematic view of the size and orientation of a small central bar approximately 3 kpc long, based on numerous observations. The location and motion of the Sun are also shown. The innermost contour shows the bar shape and orientation suggested by 2.4 μm observations of diffuse starlight. Contours further out, based on the space density of asymptotic giant branch (AGB) stars, show the bar extending out to a ring-like structure with a radius of approximately 5 kpc, which might be the densest, innermost portions of spiral arms. (b) The positions of about 5000 AGB stars observed with IRAS, showing the bar-shaped distribution of the stars. ((a) Based on data from Blitz and Spergel, 1991 and Weinberg, 1992; (b) Nikolaev and Weinberg, 1997)

The distribution of individual stars

High-metallicity red giants that are burning helium in their cores, and a group of asymptotic giant branch (AGB) stars (see Figure 1.28), have been observed towards the Galactic centre. The locations of about 5000 AGB stars are shown in Figure 1.31b, and contours of the number density of such stars are shown in Figure 1.31a. These stars have a bar-like distribution that is consistent with the 2.4 μm results included in Figure 1.31a. The distribution of AGB stars suggests the presence of three components: a non-circular central 'bulge', a bar-like structure (about 6 kpc in length) oriented at 36° to the line of sight, and a structure 5 kpc from the Galactic centre that might be inner portions of the spiral arms (the outer contour of Figure 1.31a).

■ Why don't the AGB stars show the spiral-arm structure as clearly as, say, the HII regions in Figure 1.23?

❑ The best tracers of spiral arms are short-lived objects that must still be close to the place they formed, such as O- and B-type stars. HII regions are used as tracers because they are excited by such stars. AGB stars are generally older objects that are only now reaching the ends of their lives. The AGB stars therefore cannot be expected to be good tracers of spiral arms.

Are bars unavoidable?

Once the existence of a bar is accepted, it prompts the question: 'Why does the bar exist?' Computer models of the motions of stars in galactic discs have shown that bars are very natural features. It turns out that discs without bars are difficult to produce, because a smooth disc dominated by rotational rather than random motion is unstable to the formation of a bar. That is, asymmetries in a disc can rapidly (within a few rotations of the galaxy) produce a bar structure; this behaviour of discs is called the **bar instability**. This suggests that there is no need to explain where the Galactic bar came from; the difficulty may be to explain why non-barred spirals do *not* have one.

1.5.4 The central black hole

We saw in Section 1.3 that it is possible to measure the mass of an object by its gravitational influence on other bodies that orbit it. We performed calculations for the mass of the Sun (for which we obtained the mass $M_\odot = 2 \times 10^{30}$ kg), and for the inner part of the Galaxy out to the radius of the Sun (for which we obtained $1 \times 10^{11} M_\odot$). In this section we examine another mass estimate, that for the compact object that has been found at the very centre of the Milky Way. Our findings have implications not only for the conditions in the Galactic centre, but also for the formation of our Galaxy and for many other galaxies – a topic that is discussed in subsequent chapters.

The central region of the Galaxy contains two features that are particularly prominent in radio continuum maps, named Sagittarius A (Sgr A) and Sagittarius B2 (sometimes written Sgr B_2). Sagittarius B2 is associated with a number of HII regions and is usually regarded as a star-forming region. One of the first signs that the Galactic centre is a particularly interesting location in the Galaxy was the discovery that Sagittarius A encloses an intense, unresolved radio source known as **Sagittarius A*** (pronounced 'Sagittarius A star' and often written Sgr A*); it is this latter object which is thought to mark the precise centre of the Milky Way. Within

0.04 pc (1 arcsec) of Sgr A* lies a star cluster whose members include hot, massive, and therefore young stars, whose age appears to be only 10^7 yr. This suggests that either a burst of star formation occurred that recently, or else collisions and mergers of stars have occurred in the dense stellar environment.

It is possible to estimate the mass enclosed within a certain volume by studying the motion of material around that volume, as we did in Section 1.3, provided the observed motion is the result of gravitational forces alone. In the case of the Galactic centre, observations have been performed of one particular star whose motion about the centre has been mapped over two-thirds of a complete orbit (Figure 1.32). The *velocity dispersion*, which measures the spread of orbital speeds obtained from the observations of a range of objects near the Galactic centre (Figure 1.33) is remarkably similar to the Keplerian $1/r^{1/2}$ decline (Figure 1.13b) seen in the Solar System, thus indicating that a highly concentrated mass dominates the environment.

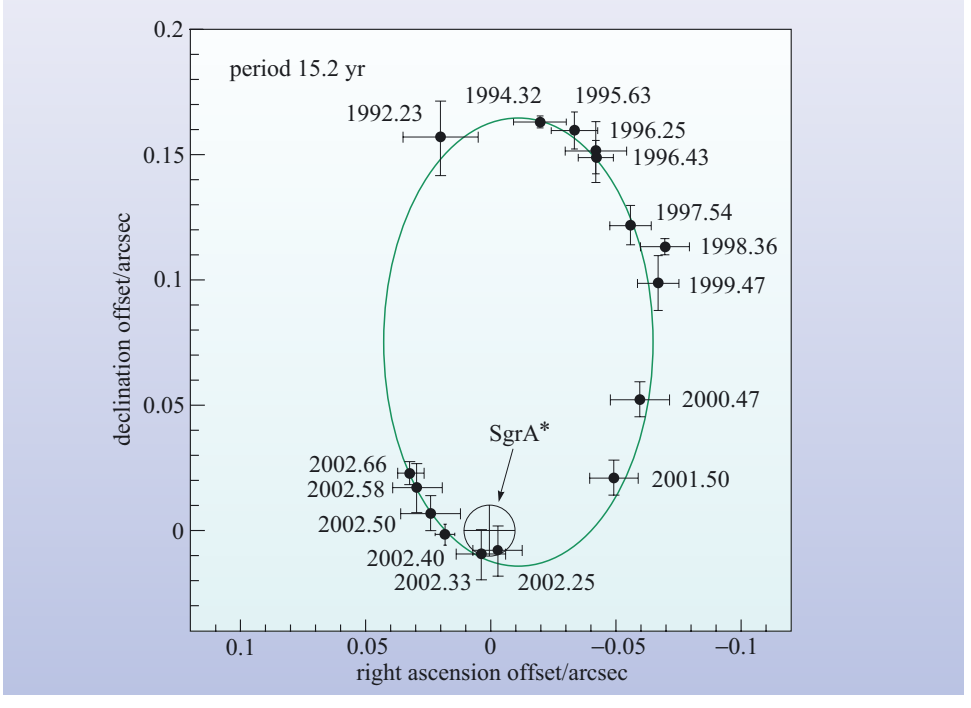


Figure 1.32 The motion of a star within the inner 0.008 pc (0.2 arcsec) of the Galaxy, observed over ten years of its 15-year orbit. (Schödel *et al.*, 2002)

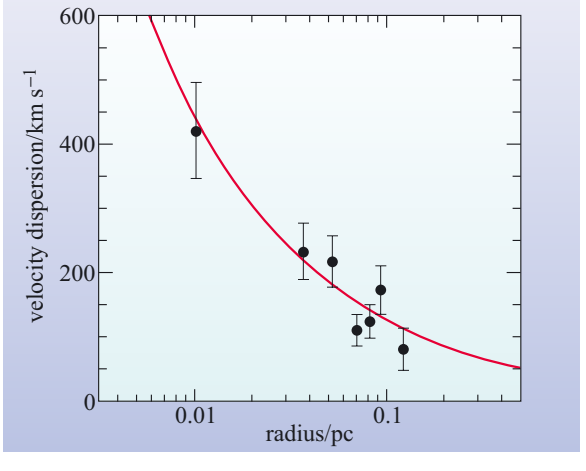


Figure 1.33 The velocity dispersion of objects near the Galactic centre, as a function of their distance from the centre. (Ghez *et al.*, 1999)

The resulting limits on the enclosed mass within spherical regions of various radii, centred on the Galactic centre, are shown in Figure 1.34. The figure indicates that a mass of about $2.6 \times 10^6 M_\odot$ is contained within 0.001 pc of the centre. A highly probable explanation is that the enclosed mass takes the form of a *massive black hole* with a mass of $2.6 \times 10^6 M_\odot$ or so. Radio measurements of Sgr A* place an upper limit on its size, indicating that its radius can't be more than a few times the distance from the Earth to the Sun. Although only an upper limit, this is already sufficient to confirm that the central object is very compact as well as enormously massive.

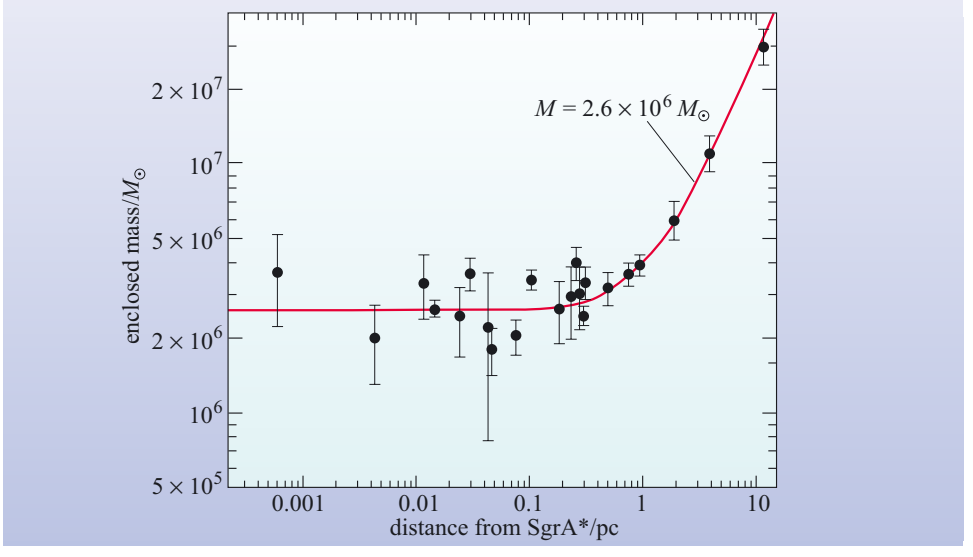


Figure 1.34 The mass $M(r)$ enclosed within spherical volumes of various radii r, centred on the Galactic centre. The enclosed mass tends to a limit $\approx 2.6 \times 10^6 M_\odot$ as r approaches zero, suggesting that this large mass is concentrated at the position of Sgr A*, perhaps as a black hole. (Schödel *et al.*, 2002)

What are the implications for the formation of the Galaxy of finding a massive black hole in its centre? You will see in Chapter 3 that black holes probably exist at the centres of many other galaxies, although those black holes cannot be studied in the same detail as the one in the centre of the Milky Way. From observations of the Milky Way and other galaxies, it appears that the presence of a massive central black hole may be connected with the existence of a spheroidal component. Much of this work is still in progress, so the picture is still emerging.

We have now completed our survey of the major components of the Galaxy, and have seen hints of the processes that link them together. The final section of this chapter brings these ideas together to look at the evolution of the Galaxy.

1.6 The formation and evolution of the Milky Way

At several points in this chapter we have seen the threads of a theory for the formation and evolution of the Galaxy. In this final section of Chapter 1 we pull these various pieces of evidence together to see what light they throw on the origin and evolution of our Galaxy.

In attempting to use observations of the Milky Way to illuminate its origin, we have to wind back the clock of star formation to see what the Galaxy would have been like at earlier times. We can hope for some success in doing this, because the Galaxy contains stars spanning a wide range of ages, and we can observe their different characteristics. Nevertheless, there is a limit to how much this process can reveal, since some information concerning the formation of the Galaxy has been erased, or

at least obscured, by subsequent events. Consequently, the observations must be united with theoretical models that work forwards in time rather than backwards. Some models, such as the model of cosmic recycling that was introduced in Section 1.2.5 and is elaborated below, consider the way that material in the ISM is processed through stars, chemically enriched and injected back into the ISM again. Other models, which are discussed in subsequent chapters, look at the way dark matter and normal matter interact to form the pool of gas from which the first Galactic stars formed. Yet another series of models looks at the way structure emerged in the early Universe, and how this allowed matter to form the kind of aggregates that developed into the groups and clusters of galaxies that surround us. As you will see later, understanding the formation and evolution of the Milky Way involves every part of this story, and every chapter in this book … and more. For that reason our objectives in this section are rather modest. Rather than trying to present a full account of the origin and evolution of the Milky Way, we content ourselves with surveying some relevant observations, raising some difficult questions, and preparing the way for wider ranging discussions of galaxy formation in later chapters.

1.6.1 The evolution of the interstellar medium

We now take a slightly more detailed look at the model for the chemical evolution of the Galaxy (i.e. cosmic recycling) that was introduced in Section 1.2.5. We begin with a schematic diagram (Figure 1.35) that sets out the flows of mass between different constituents and through different processes in the enrichment cycle.

The ISM is a key element of the evolutionary picture, because it is the birthplace of stars and it is the repository for material ejected from stars towards the end of their lives. Dense molecular clouds form from the low-density ISM, and these give rise to a new generation of stars in a star cluster. The stars go on to produce metals. The metals may be ejected back into the ISM at the end of a star's life, or they may be locked away in a dense stellar remnant such as a white dwarf, a neutron star or a

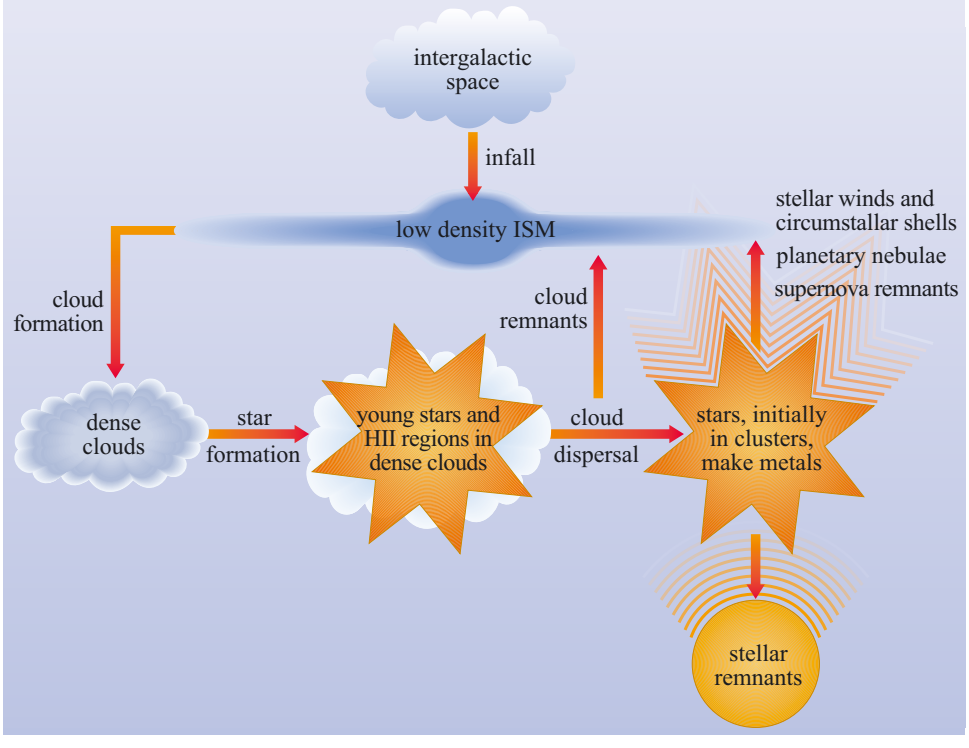


Figure 1.35 The evolution of the interstellar medium includes cycles, sources (infall from the intergalactic medium) and sinks (losses to stellar remnants).

black hole. Metals that are ejected enrich the low-density ISM, and allow the cycle of enrichment to continue.

The return of matter from stars to the ISM occurs in several ways, via:

- stellar winds from cool giants/supergiants;
- the ejection of planetary nebulae, which are the envelopes of intermediate mass stars – with initial masses less than about $11M_{\odot}$ – shed when the stars evolve to become white dwarfs;
- supernova ejecta containing most or all of a star's mass, some of which is enriched with metals.

A supernova produces an energetic shock wave that heats and ionizes a bubble of material in the ISM. Where large numbers of supernovae are occurring, such as in very active star-forming regions, these bubbles overlap, producing a very large cavity of hot, ionized gas called a **superbubble**. This hot material is restricted from expanding in the disc plane due to the pressure of the surrounding ISM, but it can expand vertically due to the much lower pressure of gas in the stellar halo. ISM material heated by supernovae may therefore enter the stellar halo via so-called **chimneys**, which are passages that open up where the hot, expanding material breaks out of the disc. Eventually, the hot material cools and probably returns in fragments to the disc. This mechanism has been termed a **Galactic fountain**, because of the motion of material squirted away from the disc and brought back again by gravity. It has been estimated that roughly $10M_{\odot}$ of material per year passes around this cycle.

Although the stellar halo now contains very little gas, almost none compared to the disc, there nevertheless exists a tenuous body of hot halo gas that is sometimes referred to as the **gaseous corona**. Figure 1.36 reveals the presence of this hot gas, which causes absorption lines in the ultraviolet (UV) spectra of distant stars. This coronal gas may be sustained by the Galactic fountain.

Is there any evidence for a returning flow of cooler gas? Possibly. Observations at wavelengths near 21 cm reveal the presence of clouds of atomic hydrogen, moving rapidly relative to the Sun. Many of these clouds have a significant velocity component *towards* the Sun. They are referred to as **high-velocity clouds**. It is difficult to determine the distance of these clouds, but they are known to be further

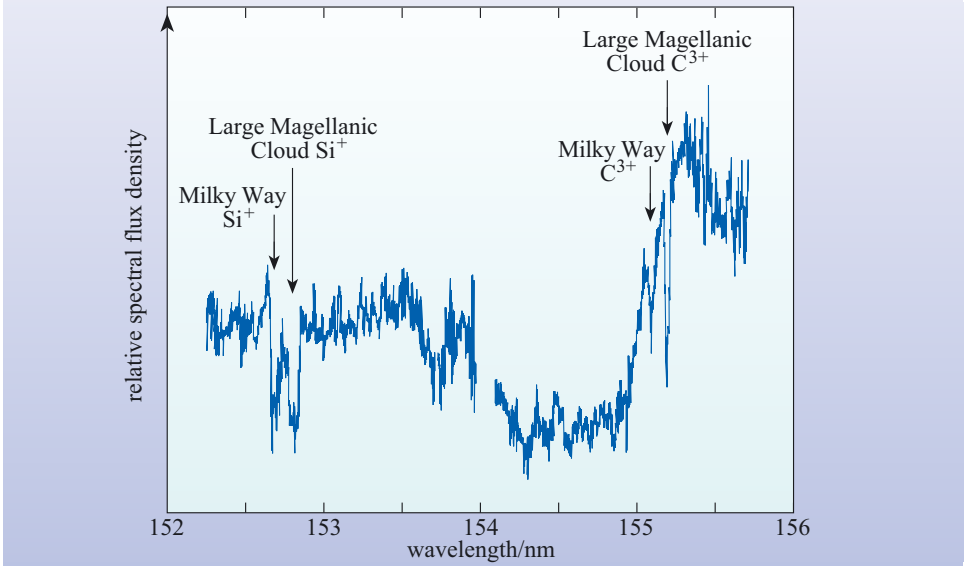


Figure 1.36 Part of the spectrum of the star HD 38282 which resides in a nearby galaxy, the Large Magellanic Cloud. Two pairs of absorption features are arrowed, one pair due to singly ionized silicon (Si^{+}), the other due to triply ionized carbon (C^{3+}). In each case, the shorter-wavelength member of the pair is due to absorption by gas in the gaseous corona of the Milky Way. The longer-wavelength member is due to similar absorption in the Large Magellanic Cloud, but is Doppler-shifted due to the relative motion of the two galaxies. (de Boer and Savage, 1982)

away than most disc stars. They are often *presumed* to be in the Galactic halo, in which case they could represent the returning flow of gas ejected by the Galactic fountain. However, it is also possible that at least some of them are much further away than that. The distances of some clouds have been estimated at 400 kpc to 1000 kpc, comparable to the 750 kpc distance of the spiral galaxy M31. This would make them intergalactic rather than merely interstellar. Their origin is still a mystery.

■ How could the speed of such clouds, towards or away from the Sun, be determined?

❑ By looking for a shift in the wavelength (or frequency) of the 21 cm line, and assuming that it is a Doppler effect due to the motion of the cloud.

In addition to the return of material to the ISM, there is also the possibility of new matter entering the enrichment cycle, due to the infall of gas and dust from outside the Galaxy. This would come from the *intergalactic* medium, which consists of low-density gas that is found between galaxies. Although little is known about such material in the neighbourhood of the Milky Way, the high-velocity clouds may be examples of such material.

Once the various sources and sinks of gas are accounted for, what is the net rate of gain or loss of ISM gas today? The rate at which matter is entering the ISM is probably between $0.4M_\odot$ and $3M_\odot$ per year, of which no more than about $1.4M_\odot$ is infall from the intergalactic medium. The present rate at which mass is leaving the ISM, to form new stars, is probably somewhere between about $3M_\odot$ and $10M_\odot$ per year, and so it is likely that the rate of loss by the ISM is still exceeding its rate of gain.

The ISM evolves not only in the fraction of mass that it contains, but also in its composition. The metallicity of a main sequence star indicates the metallicity of the ISM at the time that star formed. As noted earlier, Pop. II stars in the stellar halo have a wide range of metallicities, from $Z = 2 \times 10^{-6}$ to 0.002. Stars of progressively higher metallicity record the chemical composition of the Galaxy as it was enriched by the ejecta of successive generations of stars and supernovae. Observations of very old stars and gas clouds whose metal content has not been enriched by nucleosynthesis suggest that, when the Galaxy formed, the relative *numbers* of nuclei of H : He : metals were 93 : 7 : 0. This ratio is not very different from the present one, which is something like 91 : 9 : 0.1. Thus, the ISM still consists mainly of hydrogen and helium. However, over the lifetime of the Milky Way, the metallicity, Z, has increased from 10^{-9} to 0.04.

1.6.2 The evolution of the stellar populations

The stellar content of the Milky Way evolves in a number of ways as a result of cosmic recycling. One way is the change in the average chemical composition of stars that has taken place over time. We have already seen that within the Galaxy the youngest disc stars are also those with the highest metallicity. This makes good sense in terms of the picture of cosmic recycling. A corresponding tendency for old stars to have low metallicities was noted when we discussed the halo. Recall that the halo contains some of the oldest stars in the Galaxy and these have the lowest metallicities (down to $Z = 2 \times 10^{-6}$ in some cases). This correlation between age and metallicity is known as the **age–metallicity relation**.

However, the correlation between age and metallicity is imperfect; bulge stars are also very old, but, as we have seen, some have similar metallicities to the Sun. This fact is attributed to the higher rate of star formation and cosmic recycling in the relatively

dense bulge. But the need to recognize this limitation on the age–metallicity relation provides a useful reminder that in a complicated galaxy like the Milky Way, star formation is unlikely to have a simple evolutionary history.

Besides the increase in metallicity with time, there has also been a change in the distribution of stars. Many stars in the old components of the Galaxy – the bulge and the stellar halo – follow orbits that take them far from the Galactic plane. On the other hand, disc stars are seldom found more than a few hundred parsecs from the mid-plane. The vertical distribution of the disc stars reflects that of the gas from which they formed, and the gas now has a scale height of only 150 pc. But what was it like in the past? The thick-disc stars have a larger scale height, greater ages, and lower metallicities. Does this indicate that the disc used to be thicker, and that it has collapsed closer towards the Galactic plane, or alternatively did the thick disc used to be thin and has it been fluffed up to its current size by some energetic process, such as the collision between the Milky Way and another galaxy? And what about the stellar halo? This has an even greater extent than the disc and bulge, extending out to tens of kiloparsecs, and it contains the oldest stars. Did the Galaxy form its first stars when it was still this large, postponing the majority until later when the gas had collapsed into the Galactic plane to form the disc, or did the halo stars join the Galaxy, ready made, from another source?

These are typical of the unanswered questions relating to star formation that underlie the developing model for the formation and evolution of the Galaxy.

QUESTION 1.13

By considering the processes in Figure 1.35, describe, in a couple of sentences, the amount and composition of the ISM that will exist in the Galaxy a *long* time in the future.

1.6.3 New arrivals

To conclude our discussion of the evolution of the Galaxy, we note that some of its stars are recent arrivals. Astronomers have speculated for several decades that galaxies would on occasions collide, and that these events would lead to some remarkable shapes for the affected galaxies, as well as triggering new bursts of star formation as gas clouds are compressed in the collisions. (Stars, in contrast to gas clouds, are very small and almost never collide with one another, so the evolution of gas and stars differ greatly in collisions.) There are many examples of galaxies that appear to have collided recently; examples of such events are discussed in Chapter 2. It has even been speculated that the Galactic halo was populated by the remnants of small galaxies that collided with the Milky Way, and were captured in the process. What astronomers didn't realize until 1994 was that the Milky Way is involved in a collision *right now*.

What was discovered in 1994 is that a small galaxy, subsequently named the **Sagittarius dwarf galaxy**, is colliding with our Galaxy, and being shredded in the process (Figure 1.37). This collision is taking place on the far side of the Galaxy, which is why it went unobserved for so long. Nonetheless, the collision is adding stars to the halo. But that's not all; several globular clusters that were previously believed to be ordinary members of the Milky Way are now recognized as having been captured from the Sagittarius dwarf. These observations emphasize that even though most of the stellar halo is very old, it is still a dynamic part of the Galaxy. The Milky Way continues to evolve!

Figure 1.37 The collision of the Sagittarius dwarf galaxy with the Milky Way. The stars of the Sagittarius dwarf are very faint, so the dwarf galaxy is shown here as a red outline only. Astronomers have worked out that the Sagittarius dwarf is currently located on the far side of the Galaxy, almost in line with the Galactic centre (compare with Figure 1.1), and close to the southern side of the disc. (R. Ibata, R. Wyse and R. Sword/NASA)

1.7 Summary of Chapter 1

We conclude this chapter with a summary table of the major structural components of the Milky Way (Table 1.1), followed by a revision of key points.

Table 1.1 The major structural components of the Milky Way.

Component	Shape	Dimensions	Main forms of baryonic matter		Mass/$M_\odot$	Motion
			Stellar	Gaseous		
dark-matter halo	oblate spheroid?	>50 kpc?	?	?	$\sim 10^{12}$	?
disc	flat disc	radius ~15 kpc				
thin disc	spiral arms	thickness ~1 kpc	Pop. I $Z \sim 0.005$ to 0.04; O, B, stars in spiral arms	dense and diffuse clouds; intercloud medium, HII regions	stars $\sim 10^{11}$ gas $\sim 10^{10}$ dust $\sim 10^{8}$	circular differential rotation; confined to plane of disc
thick disc		thickness ~2 kpc	old; intermediate Pop; $Z \sim 0.004$	little or no gas	stars $\sim 10^{10}$	almost circular, scale height ~1 kpc
spheroid			Pop. II			
stellar halo	oblate spheroid $c/a \approx 0.8$	radius > 20 kpc	$Z < 0.002$	very little gas; high-velocity clouds?	stars $\sim 10^{9}$ gas negligible	elliptical orbits, often highly inclined to Galactic plane
nuclear bulge	triaxial ellipsoid (bar)	radius 3 kpc	$Z \approx 0.02$ $2.6 \times 10^{6} M_\odot$ black hole at centre		stars + gas $\sim 10^{10}$	
hot corona				tenuous hot gas		

Overview of the Milky Way

- The Milky Way – our Galaxy – is a barred spiral galaxy with four major structural components: a dark-matter halo which is only detected gravitationally, a disc, a stellar halo and a central bulge. The total mass is $\sim 10^{12} M_{\odot}$.
- The nature of the dark matter is unclear, but it may account for 90% of the total mass.
- The directly detectable matter consists mainly of stars (~90% by mass), gas (~10%) and dust (~0.1%).
- The disc is about 30 kpc in diameter and 1 kpc thick. The stellar halo is roughly spherical; its diameter is difficult to determine but estimates of more than 40 kpc are common. The nuclear bulge is a triaxial bar extending out to about 3 kpc from the centre.
- The stars of the Milky Way may be divided into a number of populations, each of which predominates in a particular region of the Galaxy. The very youngest stars are found mainly in the spiral arms. Population I stars reside in the disc. The oldest known stars, of Population II, are found mainly as individual stars of the stellar halo, and less commonly but more recognizably in globular clusters.
- The disc is in a state of differential rotation, with stars in the vicinity of the Sun taking about 2×10^8 yr to make a complete orbit of the Galactic centre.

The mass of the Milky Way

- The rotation curve that describes movement about the Galactic centre constrains the total mass of the Galaxy. Interior to the Sun's orbit, the mass is approximately $10^{11} M_{\odot}$. It is difficult to determine how much material resides beyond the Sun's orbit, and estimates for the total mass range from $4 \times 10^{11} M_{\odot}$ to $6 \times 10^{12} M_{\odot}$.

The disc

- There are about 10^{11} stars in all, with a total mass $\sim 10^{11} M_{\odot}$.
- Most stars and gas are approximately 70% hydrogen, 28% helium, and 2% heavier elements (metals) by mass.
- Hydrogen occurs in the form of molecules (H_2), atoms (HI) or ions (HII), according to local conditions. Molecular hydrogen is difficult to detect, however, so carbon monoxide (CO) is used as a tracer of H_2.
- Dust consists of μm-sized solid compounds, especially graphite and silicates with icy mantles, and accounts for about 1% of the ISM by mass.
- The (number) density of the disc's visible constituents, stars, gas and dust, falls off with distance from the mid-plane. In each case this is described by a scale height.
- The Sun is one of the Pop. I stars, located about 8.5 kpc from the centre of the Galaxy, close to the mid-plane of the disc. It is part of the thin-disc subpopulation that has a scale height of about 300 pc. There is also a thick disc subpopulation with a scale height of about 1000–1300 pc.
- The spiral arms are sites of active star formation. Attempts to trace the arms make use of young, short-lived objects in the disc such as bright HII regions, young open clusters, OB associations, dense clouds and clouds of neutral hydrogen gas.

- The spiral arm pattern might be caused by density waves – relatively slow-moving regions of density enhancement that rotate 'rigidly' around the Galactic centre. The compression of dense molecular clouds as they enter these regions might trigger the birth of the short-lived features that trace the spiral arms.

The stellar halo and bulge

- The stellar halo is roughly spherical but with polar flattening, resulting in an oblate spheroidal shape. Its radius is more than 20 kpc, although its density falls off with distance.
- The main constituents of the stellar halo are old, low-metallicity stars of Pop. II. The total mass of these stars is about $10^9 M_\odot$.
- About 1% of the halo stars are contained in globular clusters: spherical gatherings of stars, up to about 50 pc across, that contain 10^4 to 10^6 members with ages in the range (12 to 15) $\times$ 10^9 years.
- The stellar halo includes a corona of tenuous, hot ($\approx 10^6$ K) gas, thought to be heated by Galactic fountains that are powered by supernovae. The high-velocity clouds of atomic hydrogen may be located in the stellar halo or may belong to intergalactic space.
- The bulge seems to have the form of a triaxial bar extending out to 3 kpc from the Galactic centre. Its outer regions rotate at about 100 km s^{-1} and its total mass is around $10^{10} M_\odot$.
- The bulge mainly consists of old stars of the same age as the stellar halo, although their metallicities seem similar to that of the Sun's.
- Near the Galactic centre, the compact radio source Sagittarius A* lies at the heart of the Milky Way. From the motions of stars orbiting close to the Galactic centre, Sagittarius A* is believed to be a black hole of mass $2.6 \times 10^6 M_\odot$.

The formation and evolution of the Milky Way

- A process of Galactic chemical enrichment is at work, driven by cosmic recycling, in which some of the gas removed from the ISM to form stars is returned to the ISM by stellar winds, planetary nebulae, and supernovae, enriched in the heavy elements (metals) that are formed within stars.
- Observations indicate an increase in the metallicity of newly formed stars with time, but this increase has proceeded at different rates in different parts of the Galaxy. There is also evidence that the locations of stars have changed with time, but the implications of this are unclear.
- The composition of the ISM has also evolved over time, although it is still predominantly hydrogen and helium. In addition, various sources and sinks of gas are changing the total mass of gas in the Milky Way, causing the fraction of the Galaxy's mass that is composed of gas to evolve with time.
- The Sagittarius dwarf galaxy is currently merging with the Galaxy, contributing new stars and star clusters, and ensuring that evolution is a continuing process.

Questions

QUESTION 1.14

Why do high-velocity stars have lower metallicity than the Sun?

QUESTION 1.15

The Sun is 8.5 kpc from the Galactic centre, and is thought to be 4.5×10^9 years old. Use these data, and the rotation speed you can estimate from Figure 1.13c, to calculate the number of times the Sun has orbited the Galactic centre.

QUESTION 1.16

Assuming that optical views in the disc of the Milky Way are limited to a range of 5 kpc, estimate the fraction of the volume of the stellar disc that can be surveyed optically. What is the main cause of this limitation? (Assume that the thickness of the stellar disc is about 1 kpc.)

QUESTION 1.17

If the Sun was born in association with other stars, and originally formed part of an open cluster (which is not certain), why can we no longer see any evidence of that cluster?

QUESTION 1.18

How does density wave theory solve the winding dilemma?

QUESTION 1.19

Make a list of the kinds of astronomical objects that can be used in attempts to trace the spiral arms of the Milky Way.

QUESTION 1.20

What colour are the brightest stars in the Milky Way's globular clusters? Why?

QUESTION 1.21

'The Galaxy is not an unchanging body; rather, it continues to evolve.' Summarize the evidence for this assertion.